Submitted to Bernoulli

Asymptotic Power of a Likelihood Ratio Test for a Mixture of Normal Distributions

ANDREW V. CARTER[1] and DOUGLAS G. STEIGERWALD[2]

[1] *UC Santa Barbara, Department of Statistics and Applied Probability, E-mail:* carter@pstat.ucsb.edu

[2] *UC Santa Barbara, Department of Economics, E-mail:* doug@ucsb.edu

We are interested in the power of likelihood-ratio tests for a mixture of distributions. Previous research establishes the asymptotic distribution of the tests only under the null hypothesis. By representing the likelihood ratio with Hermite polynomials, we are able to derive the asymptotic distribution under a range of alternatives that are local to the null. The limit distribution depends on a function of a Gaussian process, which is maximized over the parameter space for the mean shift. We compute the asymptotic power and show that the likelihood-ratio test can have power even when the specified parameter space does not contain the population value of the mean shift. By representing the likelihood ratio as a function of Hermite polynomials we can relate the likelihood-ratio test to the underlying moments of the data and compare the test to alternative tests based only on skewness and kurtosis. We document the power gains of the likelihood-ratio test and show the range of mean shifts over which the gains are most pronounced.

Keywords: Mixture models, Likelihood ratio tests, Asymptotic approximation.

1. Introduction

A mixture of normal distributions observes

$$y_t = \beta_0 + x_t^{\mathsf{T}}\beta + \sigma\delta s_t + \sigma u_t. \tag{1}$$

The u_t are i.i.d. standard normals, and the s_t are indicators of a shift of δ standard deviations in the intercept with $\mathbb{P}\{s_t = 1\} = \pi$. The standard likelihood asymptotics for testing for the presence of a second distribution are complicated by the lack of identifiability at the null hypothesis where either π or δ are equal to 0.

The main contributions of this paper are as follows. (i) We provide a complete description of the asymptotic power for a constrained likelihood-ratio test in Gaussian mixture models. (ii) We introduce a polynomial representation of the limiting Gaussian process. The polynomial terms correspond to moments of the underlying residuals and provide a natural interpretation of the standard series approximation. (iii) We analyze the behavior of the polynomial representation as the constrained parameter space changes. This provides a link between the size of the parameter space and the influence of higher order moments on the test statistic. (iv) We compute the asymptotic power of the test and examine how the power is reduced if the population parameter, or a local neighborhood of the parameter, does not lie within the constrained parameter space.

The main technical innovations of the paper are the following. (i) To obtain the asymptotic distribution, we use a Hermite polynomial expansion. The effect of local alternatives is to shift the mean of the random variables that form the expansion. (ii) For the asymptotic power results, we do an expansion that varies over local alternatives. For the local alternatives that are most relevant empirically, in a sample of size n the test has power to detect alternatives that are $n^{-\frac{1}{2}}$ distant from the null.

We suppose $\theta = \pi\delta$ is estimated by maximizing a likelihood function $L_n(\theta, \gamma)$ over a parameter space $\Theta \times \Gamma$. For the above model, $\gamma = (\beta, \sigma^2)$. The null hypothesis of only a single regime is

$$H_0 : \theta = 0.$$

In a local neighborhood of the null hypothesis, $L_n(\theta, \gamma)$ is relatively flat: (i) with respect to δ when $\pi = \frac{1}{2}$ and (ii) with respect to both π and δ when (π, δ) are both local to 0. This causes difficulties with standard asymptotic expansions because the terms corresponding to the second derivative are identically zero.

This lack of uniqueness has ramifications for the likelihood ratio statistic. Hartigan (1985) argues that, if the parameter space for δ is the whole real line, then not only does the statistic not have a proper asymptotic distribution but that the statistic diverges to infinity with n, although at the slow rate of $\log\log n$. Dacunha-Castelle and Gassiat (1999) and Liu and Shao (2003) show that a limiting distribution of the likelihood ratio for mixture distributions can be described by characterizing the distances between distributions. Gassiat (2002) gives an expression for the behavior of the likelihood ratio under contiguous alternatives for a simple null hypothesis. Liu and Shao (2004) and Hall and Stewart (2005) specifically study the asymptotic properties of the likelihood ratio statistic in normal mixtures where the mean and variance of the null distribution are known. As we will show below, assuming the mean and variance to be known omits an important level of detail contained in this paper. Liu and Shao establish that $\Lambda_n - \log\log n$, where Λ_n is the unconstrained likelihood ratio statistic, has an asymptotic null distribution that is an extreme value type. Hall and Stewart make a direct connection between the nonstandard behavior under the null hypothesis and loss of statistical power. They show that the unconstrained likelihood ratio test is only able to detect local alternatives that are $(n^{-1}\log\log n)^{1/2}$ distant from the null.

Our approach follows Cho and White (2007) and uses a bounded parameter space, Δ to obtain a Gaussian process limit result. As noted above, for the most empirically relevant neighborhood of the null the constrained test is able to detect local alternatives that are $n^{-1/2}$ distant from the null. There are other regions of parameter space where there is less curvature of $L_n(\theta, \gamma)$, and so the constrained likelihood ratio test is only able to detect local alternatives that are $n^{-1/3}$ or $n^{-1/4}$ distant from the null. We provide a full account of these regions in our derivation of the asymptotic distributions, but concentrate our detailed analysis of asymptotic power on the most empirically relevant local alternatives.

Although the asymptotic power of the constrained test is reduced if $\delta \notin \Delta$, we show that even in this case the constrained test can be powerful. To lay the groundwork for our results, we first present the local alternatives of interest, which have a fixed value of δ. We show how to approximate the likelihood ratio in these neighborhoods. Importantly, these expressions depend on a function of the residuals that acts as a sufficient statistic for δ, and we show how to approximate this function using Hermite polynomials. With this approximation, we can capture the behavior of the likelihood ratio under local alternatives by considering shifts in the Hermite polynomials. This paves the way for our proof of the asymptotic distribution of the likelihood ratio under local alternatives.

To approximate the null distribution, and so obtain critical values, Δ must be specified. We set $\Delta \in [-a, a]$ and introduce methods that link $[-a, a]$ to the polynomial representation of the asymptotic distribution. As the magnitude of a increases, the number of terms in the polynomial expansion must increase in order to maintain a given level of approximation accuracy. As higher order terms in the expansion correspond to higher order moments of the data, increasing the parameter space increases the influence on the test statistic of moments beyond the measures of skewness and kurtosis.

The remainder of the paper is organized as follows. In Section 2 we establish the asymptotic properties of the constrained likelihood ratio statistic under local alternatives. The asymptotic distribution is a function of a Gaussian process, rather than a χ^2 process, because of boundary constraints on π.[1] Section 3 reports the asymptotic power of the test and demonstrates that the test can have substantial power even if Δ does not contain the true mean shift.

2. Asymptotic Distribution

In our mixture model, the probability of being in state $s_t = 1$ is $\pi \in [0, 1]$ and the mean shifts δ standard deviations. It will be convenient to use the parameter $\theta = \pi\delta$ so that the expected amount of shift is $\sigma\theta$, and $\theta = 0$ under the null hypothesis. Our focus is on neighborhoods of $\theta = 0$ corresponding to the boundary value $\pi = 0$ in the null parameter space. δ is restricted in the likelihood to a bounded set $\delta \in \Delta = [-a, a]$.

We follow Chen et al. (2001) and write the log-likelihood as:

$$\begin{aligned} L_n(\theta, \gamma) &= \sum_{t=1}^{n} \log[\sigma^{-1}\phi(v_t)] + \sum_{t=1}^{n} \log[1 + \pi(e^{v_t\delta - \frac{1}{2}\delta^2} - 1)] \\ &= \sum_{t=1}^{n} \log[\sigma^{-1}\phi(v_t)] + \sum_{t=1}^{n} \log[1 + \theta Z_\delta(v_t)] \end{aligned}$$

where $\phi(\cdot)$ is the standard Gaussian density function, $v_t = \frac{y_t - x_t^{\mathrm{T}}\beta}{\sigma}$ is the (standardized)

[1] Hansen (1996) studies testing when a nuisance parameter is not identified under the null hypothesis, but in his framework the nuisance parameter is not bounded and so the limit distribution is a function of a χ^2 process.

residual under the null hypothesis, and

$$Z_\delta(v_t) := \frac{1}{\delta}(e^{v_t\delta - \frac{1}{2}\delta^2} - 1). \tag{2}$$

The function $Z_\delta(\cdot)$, through which δ enters the log-likelihood, is akin to a sufficient statistic for δ.

If $\theta = 0$, then the MLE is the OLSE with estimated residual $\widehat{v}_t = \frac{y_t - x_t^T b}{s}$. The (constrained) test statistic is the likelihood ratio

$$Q_n = 2\left[\max_{\theta\in\Theta,\gamma} L_n(\theta,\gamma) - L_n(0,b,s)\right] \tag{3}$$

where $\Theta = \Delta \times [0,1]$ with $\Delta = [-a,a]$.

Cho and White (2007) make clear that the complete asymptotic null distribution of the test statistic depends on the behavior of Q_n in local neighborhoods of three regions: fixed δ with π local to zero, fixed $\pi \neq \frac{1}{2}$ with δ local to zero, and $\pi = \frac{1}{2}$ with δ local to zero. The most complex behavior arises in the neighborhood with δ fixed and π local to zero. This is also the most important neighborhood empirically, as it is the behavior of Q_n here that allows one to distinguish between a small group of outliers (arising under the null) and the presence of a second, widely separated component that occurs infrequently. To analyze power we focus on this neighborhood with the local-to-null reparameterization $\theta_n = \frac{h}{\sqrt{n}}$. The null hypothesis is $H_0 : \theta = 0$ with alternative

$$H_{1,n} : \theta_n = \frac{h_*}{\sqrt{n}}, \quad \delta = \delta_* \tag{4}$$

implying that $\pi_n = n^{-1/2}\, h_*/\delta_*$.

Recent work in Kasahara et al. (2014) constructs a modified likelihood-ratio test with a simpler asymptotic null distribution, but the simplification comes at the expense of ignoring the local neighborhood of $\pi = 0$.

To obtain the behavior of Q_n under $H_{1,n}$, we first analyze the likelihood ratio for fixed values of δ. Let the likelihood ratio statistic for a fixed value of δ be

$$q_\delta(\beta,\sigma^2,\theta) = n\log(\frac{s^2}{\sigma^2}) + \sum_{t=1}^{n}(1 - v_t^2) + 2\sum_{t=1}^{n}\log[1 + \theta Z_\delta(v_t)],$$

which, within a $\frac{1}{\sqrt{n}}$ neighborhood of 0 is approximated by

$$q_\delta(\beta,\sigma^2,\theta) = n\log(\frac{s^2}{\sigma^2}) + \sum_{t=1}^{n}(1 - v_t^2) + 2\theta\sum_{t=1}^{n}Z_\delta(v_t) - \theta^2\sum_{t=1}^{n}Z_\delta(v_t)^2.$$

Define the sample moments $m_1 = \frac{1}{n}\sum_{t=1}^{n} Z_\delta(\widehat{v}_t)$, $m_2 = \frac{1}{n}\sum_{t=1}^{n} Z_\delta(\widehat{v}_t)^2$, and $m_c = \frac{1}{n}\sum_{t=1}^{n} \widehat{v}_t Z_\delta(\widehat{v}_t)$.

Lemma 1. *For each fixed value of δ we have that under both H_0 and $H_{1,n}$:*

$$q_\delta(\hat{\beta}, \hat{\sigma}^2, \hat{\theta}) = \frac{n\, m_1^2 \{m_1\delta > 0\}}{m_2 - 1 - \frac{\delta^2 m_c^2}{2}} + o_P(1).$$

The proof is in Appendix A.1.

Under $H_{1,n}$ the estimator $\widehat{\theta}_n$ is $n^{\frac{1}{2}}$-consistent.[2] The behavior of $q_\delta(\cdot)$ is determined by the behavior of $Z_\delta(\widehat{v}_t)$. A key to our results is the representation of $Z_\delta(\widehat{v}_t)$ as an expansion of Hermite polynomials $\{H_j(\cdot)\}_{j=1}^J$. The polynomials are defined in Appendix B, where we also show that, because the underlying data is normal, the polynomials form an orthogonal series and each term corresponds to a moment of the data.

Under standard conditions on the design variables x_t, the distribution of the residuals $\widehat{v}_t$ will converge to that of the $w_t = u_t + s_t\delta$.

Condition 1. *The covariates x_t follow a distribution such that the x_t are independent of the u_t and s_t, and the impact of the residuals on the least squares estimate are asymptotically negligible, $\left[\mathbf{X}^\top \mathbf{X}\right]^{-1} \mathbf{X}^\top \mathbf{w} = O_P(n^{-1/2})$.*

Then we can bound the error in the Hermite polynomial under both the null and under a sequence of local alternatives.

Lemma 2. *Under Condition 1 and the distribution where $\pi\delta_* = \theta_* = h_* n^{-1/2}$, for any $\epsilon > 0$*

$$\lim_{J\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{\delta\in[-a,a]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(Z_\delta(\widehat{v}_t) - \left[\sum_{j=1}^{J} \frac{\delta^{j-1}}{j!} H_j(\widehat{v}_t) \right] \right) \right| > \epsilon \right\} = 0.$$

The proof is in Appendix A.2.

This lemma allows an essentially Taylor or basis expansion of the likelihood function. One benefit is that it means uniform results over a range of δ can be derived from the limiting behavior of a finite number of random coefficients. A multivariate central limit theorem will then be sufficient to describe the asymptotic distribution of $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_\delta(\widehat{v}_t)$ under both the null and a sequence of local alternatives.

Lemma 3. *For each fixed value of J,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{j=1}^{J} \frac{\delta^{j-1}}{j!} H_j(\widehat{v}_t) \rightsquigarrow \sum_{j=3}^{J} \frac{\delta^{j-1}}{\sqrt{j!}} \left(\zeta_j + \frac{h_* \delta_*^{j-1}}{\sqrt{j!}} \right)$$

where ζ_j are independent standard normals and δ_* and h_* are the true values of those parameters.

[2]Under the other local alternatives, which we do not consider as they are less empirically relevant, $\widehat{\theta}_n$ converges at a slower rate. The Appendix contains a brief description.

The proof is in Appendix A.3. Observe that the first two terms drop out of the limiting sum because the mean and variance are estimated. If one assumed that the mean and variance were known, as in the work described in the introduction, then the first two terms would not drop out. Assuming the mean and variance are known implies that the leading term is H_1 and so the sample mean would be a reasonable test of the null hypothesis. Further, the local maximum at $\pi = 1/2$, which occurs because H_3 is zero at this point, does not appear because H_1 is not zero.

Clearly, Lemma 3 describes a Gaussian process with a particular mean and covariances, but this characterization in terms of a polynomial expansion in δ is the most useful. Critical values and asymptotic power will be calculated by simulating from a process constructed in this way.

The relationship between this process and the likelihood ratio starts with the Gaussian process

$$\mathcal{G}(\delta) = \left[\sum_{j=3}^{\infty} \frac{\delta^{j-1}}{\sqrt{j!}} \left(\zeta_j + \frac{h_* \, \delta_*^{j-1}}{\sqrt{j!}}\right)\right] \left[\delta^{-2}\left(\exp\left(\delta^2\right) - 1 - \delta^2 - \frac{\delta^4}{2}\right)\right]^{-1/2}. \tag{5}$$

The distribution of this process depends on the true value of the parameters h_* and δ_*, and it is defined as a function indexed by the parameter $\delta \in \Delta$. This definition is not well defined at $\delta = 0$ so we will extend it by continuity to $\mathcal{G}(0) = \zeta_3 + h_* \delta_*^2/\sqrt{6}$.

There are two ways in which the likelihood ratio $q_{n,\delta}$ differs from the maximum of $\mathcal{G}^2$. First, the requirement that the probability $\pi \geq 0$ implies that $\hat{\theta} = \hat{\pi}\hat{\delta}$ has the same sign as $\hat{\delta}$ and, in Appendix A.1 equation (10), we have $\hat{\theta} = m_1/(m_2 - 1 - \delta^2 m_c/2)$ so that the sign of $\hat{\theta}$ is the same as the sign of m_1. Thus, if $m_1\delta < 0$ then the MLE is $\hat{\theta} = 0$ and $q_\delta = 0$. Therefore, we only consider δ for which $\mathcal{G}(\delta)$ has the same sign as δ. Second, the actual likelihood ratio is 0 at $\delta = 0$, and the behavior of the likelihood for very small δ's is complicated because then θ could be small even if the probability π is not. The continuous extension of $\mathcal{G}(\delta)$ solves some of this issue, but as in Cho and White there is a separate local maximum in a neighborhood of $\pi = 1/2$ and $\delta = 0$ which requires inspection of the fourth-order polynomial $n^{-1/2} \sum_t H_4(\widehat{v}_t) \approx \sqrt{24}\zeta_4 + h_* \delta_*^3$.

Combining the results from Lemmas 1-3, we obtain the asymptotic distribution of the constrained likelihood-ratio test under local alternatives.

Theorem 1. *Under Condition 1 and the distribution where* $\pi\delta_* = \theta_* = h_* n^{-1/2}$,

$$Q_n \rightsquigarrow \max\left([G_+]^2, \sup_{\delta\in\Delta} \left[\mathcal{G}^+(\delta)\right]^2\right)$$

where

$$\mathcal{G}^+(\delta) = \begin{cases} \mathcal{G}(\delta)_+, & \delta \geq 0 \\ \mathcal{G}(\delta)_-, & \delta < 0. \end{cases}$$

and $G = \zeta_4 + h_* \delta_{*\,3}/\sqrt{24}$ *with* $G_+ = max(G, 0)$.

The proof is in Appendix A.4.

Derivation of the asymptotic null distribution improves on the earlier bounds on the distribution developed for χ^2 processes in Davies (1987), which builds on the bounds for Gaussian processes in Davies (1977).

3. Asymptotic Power

To compute the Gaussian process in the limit distribution, we employ the construction

$$\mathcal{G}(\delta) = \left[\sum_{j=3}^{J} \frac{\delta^{j-1}}{\sqrt{j!}} \left(\zeta_j + \frac{h_* \delta_*^{j-1}}{\sqrt{j!}} \right) \right] \left[\delta^{-2} \left(\exp(\delta^2) - 1 - \delta^2 - \frac{\delta^4}{2} \right) \right]^{-1/2}, \qquad (6)$$

where $J = 150$ following Carter and Steigerwald (2013).

Under the null hypothesis, the mean of each term in the construction equals 0, so $\mathbb{E}[\mathcal{G}(\delta)] = 0$. To gain further understanding of the null behavior of $\mathcal{G}(\delta)$, in Figure 1 we present five random draws from the process. The process is constructed for a fine grid of values, the grid mesh is 0.01, on $\Delta = [-1, 1]$.

Several features are immediate from the sample paths that contrast this Gaussian process with the more commonly known Brownian motion Gaussian process. First, the variation in $\mathcal{G}(\delta)$ does not increase with the magnitude of the argument δ. This occurs because the marginal distribution of the process is $\mathcal{N}(0, 1)$ for each value of δ, in contrast to Brownian motion where the variance of the marginal distribution increases with the magnitude of the argument.

Second, $\mathcal{G}(\delta)$ is quite smooth. This occurs because the elements of the process are strongly positively correlated, in contrast to Brownian motion which has uncorrelated increments. One further interesting feature of the limiting process is the behavior at $\delta = 0$. At this point $\mathcal{G}(\delta) = \zeta_3$ because ζ_3 is the limit of $\mathcal{G}(\delta)$ as δ approaches 0.

Critical values are obtained by Monte Carlo simulation of $\{\zeta_j\}$. For each parameter set $\Delta = [-a, a]$, we construct 1,000,000 simulations to provide precision in computing the quantiles. The critical values, for test sizes of 5% and 1% are contained in Table 1.

Table 1. Critical Values

Δ	$[-1, 1]$	$[-2, 2]$	$[-3, 3]$	$[-4, 4]$	$[-5, 5]$
5% Critical Value	4.91	5.44	6.05	6.56	6.97
1% Critical Value	7.92	8.54	9.18	9.74	10.18

Although the marginal distribution of $\mathcal{G}(\delta)$ does not change with δ, the critical values of the asymptotic null distribution do increase with a. This arises simply because as Δ expands, we are searching for a maximum of the process over a larger set. Indeed, it is this observation that leads to the need to specify a parameter set for δ, as without a set the maximum of the process is not bounded.

Under the alternative hypothesis, $\theta_n = \frac{h_*}{\sqrt{n}}$ with δ fixed at δ_* for all n. Because θ_n lies in a local neighborhood of 0, larger values of δ_* are paired with smaller values of

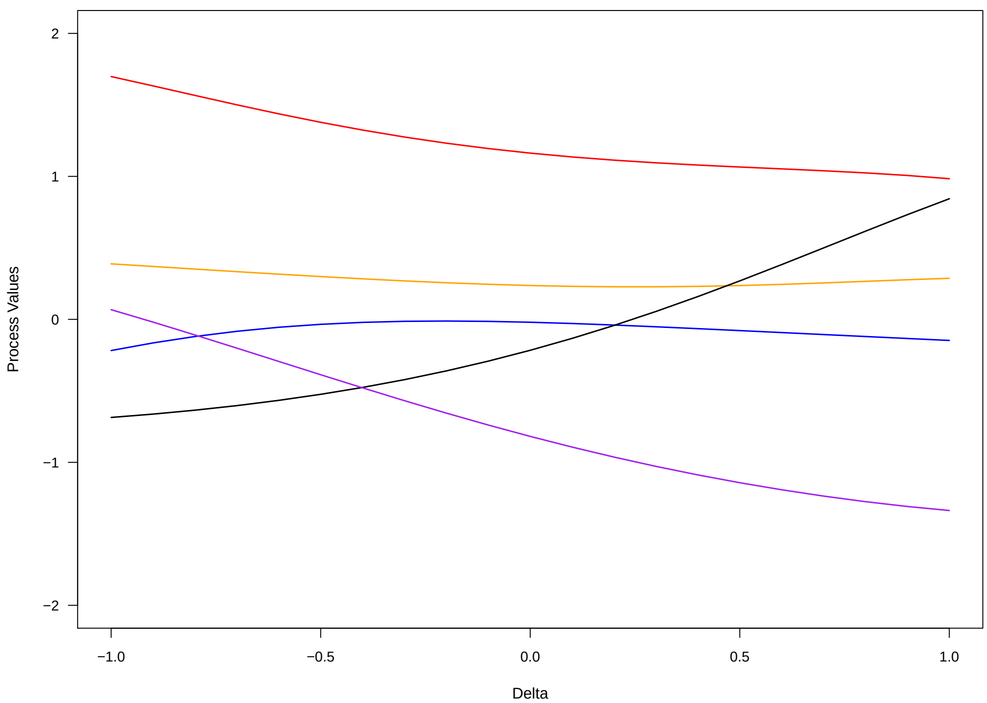

Figure 1: Sample Paths of $\mathcal{G}(\delta)$ under the Null

$\pi_n = \frac{h_*}{\delta_*} \cdot \frac{1}{\sqrt{n}}$, so that when we examine an alternative in which the shift δ_* has increased, this alternative has a smaller probability associated with the shifted regime.

The mean of each normal random variable in the construction (5) includes the component $\frac{h_*(\delta_*)^{j-1}}{\sqrt{j!}}$, so that the mean of $\mathcal{G}(\delta)$ is no longer zero. Now $\mathbb{E}[\mathcal{G}(\delta)]$ is maximized at δ_*. How does this maximum change with δ_*? Because we are in a local neighborhood, larger values of the mean separation of the second component are paired with smaller probabilities of drawing from the second component. Thus increasing δ_* could, in principle, leave the maximum of $\mathbb{E}[\mathcal{G}(\delta)]$ unchanged. Figure 2 addresses this question by graphing $\mathbb{E}[\mathcal{G}(\delta)]$ for three different values of δ_*. (Because $\mathbb{E}[\mathcal{G}(\delta)]$ is an even function, the graph is symmetric for negative values of δ.) When comparing the maximum for the three values of δ_*, increasing the mean separation of the second component increases the maximum of the curve (of course, only the relative heights of the curves matter, because the magnitude of the peak depends on the arbitrary parameter h_* that is here set to 1). Interestingly, the curves lie well above the origin for values of δ less than 1 when δ_* exceeds 1. This indicates that the test should have asymptotic power even if δ_* lies outside the constrained parameter space Δ.

Table 2 contains the asymptotic power as a function of the departure from the null, δ_*, and the size of the parameter space Δ. The main entries are the rejection probabilities based on 100,000 simulations, below each main entry is the Monte Carlo standard error in parentheses.[3]

Table 2. Asymptotic Power

Δ	$\delta_* = 1$	$\delta_* = 1.5$	$\delta_* = 2$
		5% Size	
$[-1, 1]$	0.08	0.22	0.69
	(0.0008)	(0.0014)	(0.0013)
$[-2, 2]$	0.08	0.23	0.85
	(0.0007)	(0.0012)	(0.0010)
$[-5, 5]$	0.07	0.18	0.80
	(0.0007)	(0.0013)	(0.0013)
		1% Size	
$[-1, 1]$	0.02	0.08	0.44
	(0.0005)	(0.0008)	(0.0016)
$[-2, 2]$	0.02	0.08	0.66
	(0.0004)	(0.0009)	(0.0016)
$[-5, 5]$	0.02	0.06	0.59
	(0.0004)	(0.0007)	(0.0015)

Studying each row in the table reveals that the asymptotic power increases with the magnitude of the mean separation of the second component. This quantifies the pattern in Figure 2 that shows an increase in asymptotic power for local alternatives with a greater mean shift but lower probability. Comparing the entries within each panel column shows that the power does not necessarily change monotonically as the parameter space Δ is enlarged. If Δ does not contain δ_*, then enlarging Δ so that δ_* is contained within it leads to an increase in power as the maximum of $\mathbb{E}[\mathcal{G}(\delta)]$ increases. Further enlarging Δ

[3]Computation of the Monte Carlo standard errors is detailed in Appendix C.

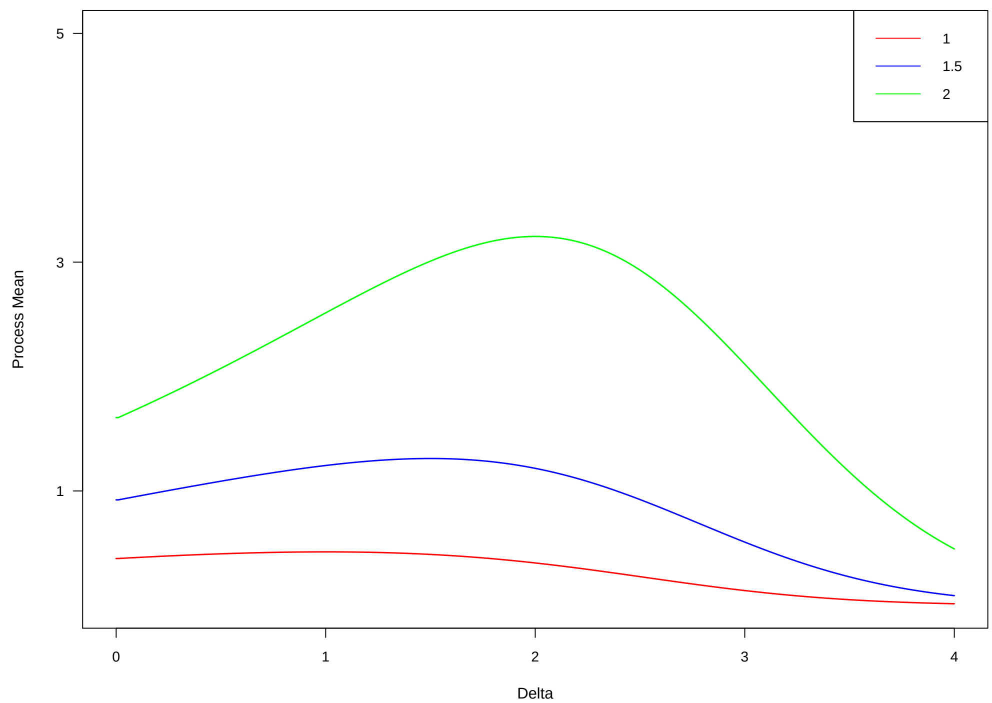

Figure 2: $\mathbb{E}[\mathcal{G}(\delta)]$ under Alternative Hypotheses

leads to a decrease in power, as the maximum of $\mathbb{E}[\mathcal{G}(\delta)]$ is unchanged and the critical values increase.

Given that there is a cost to selecting a parameter space that is too large, in the sense that δ_* lies well into the interior of Δ, it is natural to ask if constrained test based on Δ too small can deliver asymptotic power. Consider a test based on $\Delta = [-1, 1]$. What fraction of the available asymptotic power, that is the maximal power of a test where $\delta_* \in \Delta$, does this test achieve? If $\delta_* = 1.5$, then the fraction of the power achieved by the test based on $[-1, 1]$ is the ratio of two entries in column 2 of Table 2: $\frac{0.22}{0.23} = 96\%$.

Table 3. Fraction of Asymptotic Power for the LR Test Based on the Constraint [-1,1]

δ_*	1.0	1.5	2.0
5% Size	100%	96%	81%
1% Size	100%	100%	66%

Table 3 shows that when the second component mean shift is constrained to be no larger than 1 standard deviation, the test still has substantial asymptotic power to detect components that are shifted by more than one standard deviation. If the second component has a mean shift of two standard deviations, then imposing the binding constraint that the mean shift is not allowed to be larger than 1 standard deviation clearly reduces power, but the constrained test captures 81% of the maximal power. As $delta_*$ grows, the reduction in power from imposing the constraint becomes more pronounced. is also interesting to note that much of the asymptotic power can be captured by a test in which the boundary of Δ is binding. This reflects the observation from Figure 2 that the shift to $\mathcal{G}(\delta)$ is pronounced at values of δ near zero. The important lesson is that δ_* can lie well within the interior, or outside of Δ, and the test can still deliver substantial asymptotic power.

4. Remarks

This paper develops the asymptotic behavior of a constrained likelihood-ratio test for a mixture of distributions under local alternatives. In so doing it gives the first analytic representation of the power of such tests, and provides a pathway to understanding the implications of the choice of the constraint (the parameter set Δ), while specifically relating the test to more commonly used tests based on the third and fourth moments.

The sequence of local alternatives is chosen to correspond to the empirically important case in which one must decide if a collection of outliers originate from a second distribution. Our analysis uses Hermite polynomials to approximate the behavior of the likelihood-ratio. These polynomials describe the asymptotic behavior of the test and illuminate how the choice of the constraint (the parameter space Δ that contains the values of the mean shift) affects the asymptotic power. The polynomials also allow for direct comparison of the likelihood-ratio test to tests that rely only on skewness and kurtosis. As the mean shift grows, indeed, if the mean shift exceeds 1 standard deviation, then the constrained likelihood-ratio test has substantially higher asymptotic power than the test based only on skewness and kurtosis.

Appendix A: Proof of Results

A.1. Proof of Lemma 1

For a fixed δ, we can use

$$q_\delta(\beta,\sigma^2,\theta) = n\log\frac{s^2}{\sigma^2} + \sum_{t=1}^n (1-v_t^2) + 2\theta\sum_{t=1}^n Z_\delta(v_t) - \theta^2\sum_{t=1}^n Z_\delta(v_t)^2.$$

as an $O_P(n\theta^3)$ approximation to the likelihood when $\theta = O_P(n^{-1/2})$.

In order to separate out the estimation of β and σ^2, we want to write the expression for q_δ around the null hypothesis residuals $\widehat{v}_t$.

$$v_t - \widehat{v}_t = \left(\frac{s}{\sigma} - 1\right)\widehat{v}_t + \frac{1}{\sigma}x_t^\mathsf{T}(b-\beta)$$

This makes the second term in q_δ

$$\sum_{t=1}^n (1-v_t^2) = n\left(1-\frac{s^2}{\sigma^2}\right) - \frac{1}{\sigma^2}\sum_{t=1}^n (b-\beta)^\mathsf{T} x_t x_t^\mathsf{T}(b-\beta). \tag{7}$$

For the next two terms in q_δ, we will use the Taylor expansion

$$\begin{aligned} Z_\delta(v_t) &= Z_\delta(\widehat{v}_t) + (v_t-\widehat{v}_t)(1+\delta Z_\delta(\widehat{v}_t)) + O([v_t-\widehat{v}_t]^2) \\ &= Z_\delta(\widehat{v}_t)[1+\delta(\frac{s}{\sigma}-1)\widehat{v}_t + \frac{\delta}{\sigma}x_t^\mathsf{T}(b-\beta)] + (\frac{s}{\sigma}-1)\widehat{v}_t + \frac{1}{\sigma}x_t^\mathsf{T}(b-\beta) + O([v_t-\widehat{v}_t]^2) \end{aligned}$$

If we know that $b-\beta = O_P(n^{-1/2})$, then the largest terms in the sum of Z_δ are

$$\begin{aligned} \sum_{t=1}^n Z_\delta(v_t) &= \sum_{t=1}^n Z_\delta(\widehat{v}_t)\left[1+\delta\left(\frac{s}{\sigma}-1\right)\widehat{v}_t + \frac{\delta}{\sigma}x_t^\mathsf{T}(b-\beta)\right] + \frac{1}{\sigma}\sum_{t=1}^n x_t^\mathsf{T}(b-\beta) + O\left(n\,[v_t-\widehat{v}_t]^2\right) \\ &= \sum_{t=1}^n Z_\delta(\widehat{v}_t) + \delta\left(\frac{s}{\sigma}-1\right)\sum_{t=1}^n \widehat{v}_t Z_\delta(\widehat{v}_t) + \frac{1}{\sigma}\sum_{t=1}^n x_t^\mathsf{T}(b-\beta) + O_P(1) \\ &= nm_1 + n\delta\left(\frac{s}{\sigma}-1\right)m_c + \frac{1}{\sigma}\sum_{t=1}^n x_t^\mathsf{T}(b-\beta) + O_P(1) \end{aligned}$$

and

$$\sum_{t=1}^n Z_\delta(v_t)^2 = \sum_{t=1}^n Z_\delta(\widehat{v}_t)^2 + O_P(\sqrt{n}) = nm_2 + O_P(\sqrt{n})$$

This gives us the approximation of q

$$q_\delta(\beta,\sigma^2,\theta) = n\left[\log\frac{s^2}{\sigma^2}+1-\frac{s^2}{\sigma^2}\right]-\frac{1}{\sigma^2}\sum_{t=1}^n (b-\beta)^\mathsf{T} x_t x_t^\mathsf{T}(b-\beta)+ \\ +2n\theta m_1 + 2n\theta\delta\left(\frac{s}{\sigma}-1\right)m_c + \frac{2\theta}{\sigma}\sum_{t=1}^n x_t^\mathsf{T}(b-\beta) - n\theta^2 m_2 + o_P(1).$$

This q_δ is maximized at $\hat\beta$ and $\hat\sigma$

$$b-\hat\beta = \theta\hat\sigma\left[X^\mathsf{T}X\right]^{-1}\sum_{t=1}^n x_t \tag{8}$$

$$\frac{s^2}{\hat\sigma^2} = 1+\delta\theta m_c + O_P(1). \tag{9}$$

These estimates confirm that $b-\hat\beta$ and $s^2/\hat\sigma^2-1$ are small ($O_P(\theta)$).

Plugging in these quantities to the formula for q_δ, we get

$$\begin{aligned} q_\delta(\hat\mu,\hat\sigma^2,\theta) &= 2n\theta m_1 - n\frac{\delta^2\theta^2 m_c^2}{2} + 2n\delta\theta m_c\left(\sqrt{1+\delta\theta m_c}-1\right) - n\theta^2(m_2-1) \\ &= 2n\theta m_1 - n\theta^2\left(m_2-1-\frac{\delta^2 m_c^2}{2}\right)+o_P(1) \end{aligned}$$

Thus, our estimate of θ will be

$$\hat\theta = \frac{m_1}{m_2-1-\frac{\delta^2 m_c^2}{2}}. \tag{10}$$

We need to insure that this estimator is inside our parameter space with $\pi \geq 0$ implying that the sign of θ and δ must be the same. Therefore, if $m_1\delta < 0$, then $\hat\theta$ is actually taken to be 0.

It then follows that the test statistic at δ is

$$q_\delta(\hat\mu,\hat\sigma^2,\hat\theta) = \frac{n\ m_1^2}{m_2-1-\frac{\delta^2 m_c^2}{2}}\{m_1\delta>0\}+o_P(1)$$

as required by our lemma.

A.2. Proof of Lemma 2

The generating function expansion of Z_δ implies that

$$Z_\delta(\widehat{v}_t) - \sum_{j=1}^J \frac{\delta^{j-1}}{j!}H_j(\widehat{v}_t) = \sum_{j>J}\frac{\delta^{j-1}}{j!}H_j(\widehat{v}_t).$$

We begin by bounding the error term

$$\mathbb{P}\left\{\sup_{\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j>J}H_j(\widehat{v}_t)\frac{\delta^{j-1}}{j!}\right|>\epsilon\right\}\leq\mathbb{P}\left\{\sup_{\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j>J}H_j(u_t+s_t\delta)\frac{\delta^{j-1}}{j!}\right|>\epsilon\right\}+$$
$$+\mathbb{P}\left\{\sup_{\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j>J}\left[H_j(\widehat{v}_t)-H_j(u_t+s_t\delta)\right]\frac{\delta^{j-1}}{j!}\right|>\epsilon\right\}$$

The second term above is shown to be asymptotically negligible by Lemma 3.1 in Carter and Steigerwald (2018) under Condition 1.

We will employ Chebyshev's inequality to bound the first probability. It will be simpler if we denote $w_t = u_t + s_t\delta_*$. We need to bound

$$\mathbb{E}\sup_{\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j>J}H_j(w_t)\frac{\delta^{j-1}}{j!}\right|^2$$

The function is increasing in δ and so it will reach it's maximum at the boundary $\delta = a$. The v_t are independent so that the expectation is bounded using

$$\frac{1}{n}\sum_{k>J}\sum_{j>J}\frac{a^{j+k-2}}{j!\,k!}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathbb{E}H_j(w_t)H_k(w_s)$$
$$=\sum_{k>J}\sum_{j>J}\frac{a^{j+k-2}}{j!\,k!}\operatorname{Cov}(H_j(w_t),H_k(w_t))+n\left[\sum_{j>J}\frac{a^{j-1}}{j!}\mathbb{E}H_j(w_t)\right]^2 \quad (11)$$

By Lemma B.1,

$$\sum_{k>J}\sum_{j>J}\frac{a^{j+k-2}}{j!\,k!}\left[\operatorname{Cov}(H_j(w_t),H_k(w_t))+n\mathbb{E}H_j(w_t)\mathbb{E}H_k(w_t)\right]=$$
$$=\sum_{j>J}\frac{a^{2j-2}}{(j!)^2}\operatorname{Var}(H_j(w_t))+2\sum_{j>J}\sum_{k>j}\frac{a^{j+k-2}}{j!\,k!}\operatorname{Cov}(H_j(w_t),H_k(w_t))+n\left[\sum_{j>J}\frac{a^{j-1}}{j!}\mathbb{E}H_j(w_t)\right]^2$$
$$\leq\sum_{j>J}\frac{a^{2j-2}}{j!}+n\left[\sum_{j>J}\frac{a^{j-1}}{j!}\frac{h_*\delta_*^j}{\sqrt{n}}\right]^2+O(n^{-1/2})$$
$$\to\sum_{j>J}\frac{a^{2j-2}}{j!}+h_*^2\delta_*^2\left(\sum_{j>J}\frac{(a\delta_*)^{j-1}}{j!}\right)^2 \quad (12)$$

as $n\to\infty$.

By the properties of our exponential Taylor expansion, these sums are bounded

$$\sum_{j>J} \frac{a^{2j-2}}{j!} \leq \mathsf{e}^{a^2} \frac{a^{2J}}{J!}$$
$$\sum_{j>J} \frac{(a\delta_*)^{j-1}}{j!} \leq \mathsf{e}^{a\delta_*} \frac{a^J \delta_*^J}{J!}$$

which goes to 0 as $J \to \infty$.

Therefore, the expected value of the squared random variable goes to 0, and then Chebyshev's will imply that the probability goes to zero for any $\epsilon > 0$.

A.3. Proof of Lemma 3

Following the proof of Lemma 2,

$$\sqrt{n} m_1 \approx Q_J(\delta) = \sum_{k=1}^{J} \frac{\delta^{k-1}}{k!} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} H_k(\widehat{v}_t) \right] \tag{13}$$

From the expansion in (13), we apply a multivariate central limit theorem argument to the J averages of the Hermite polynomial terms. This requires establishing the moments of $H_k(\widehat{v}_t)$.

By definition, the $\widehat{v}_t$ are standardized using the OLS estimators and thus

$$\sum_{t=1}^{n} H_1(\widehat{v}_t) = \sum_{t=1}^{n} \widehat{v}_t = 0$$
$$\sum_{t=1}^{n} H_2(\widehat{v}_t) = \sum_{t=1}^{n} \widehat{v}_t^2 - n = 0$$

so we will be concerned only with $j \geq 3$.

We can approximate the behaviour of this function of the $\widehat{v}_t$ by

$$\sum_{j=3}^{J} \frac{\delta^{j-1}}{j!} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} H_j(\widehat{v}_t) \right] \approx \sum_{j=3}^{J} \frac{\delta^{j-1}}{j!} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} H_j(u_t + \delta_* s_t) \right].$$

This follows from Lemma 3.1 in Carter and Steigerwald (2018).

In Appendix B, the properties of random variables $H_j(x)$ are explored. Under the null hypothesis, where the error term u_t is a standard normal

$$\mathbb{E} H_j(u_t) = 0, \qquad \mathrm{Var}(H_j(u_t)) = k!,$$

and orthogonality of the polynomials implies $\mathrm{Cov}(H_k(u_t), H_j(u_t)) = 0$. Under the mixture model alternative where $\mathbb{E}(u_t + \delta_* s_t) = \theta_* = h_* n^{-1/2}$, Lemma B.1 states that

$$\mathbb{E} H_j(u_t + \delta_* s_t) = \frac{h_* \delta_*^{j-1}}{\sqrt{n}}$$

while the variances and covariances are within $O(n^{-1/2})$ of those from the null hypothesis. Therefore, the joint distribution of

$$\left\{ \frac{1}{\sqrt{nk!}} \sum_{t=1}^{n} H_k(u_t + \delta_* s_t) \right\}_{3 \le k \le J} \rightsquigarrow \mathcal{N}\left(\left\{ \frac{h_* \delta_*^{k-1}}{\sqrt{k!}} \right\}_{3 \le k \le J}, \mathbf{I}_{J-2} \right). \tag{14}$$

This implies that linear functions of these coordinates also have a limiting normal distribution,

$$\sum_{j=1}^{J} \frac{\delta^{j-1}}{\sqrt{j!}} \left[\frac{1}{\sqrt{nj!}} \sum_{t=1}^{n} H_j(\widehat{v}_t) \right] \rightsquigarrow \sum_{j=3}^{J} \frac{\delta^{j-1}}{\sqrt{j!}} \left[\zeta_j + \frac{h_* \delta_*^{j-1}}{\sqrt{j!}} \right],$$

and thus the desired result is established.

A.4. Proof of Theorem 1

To connect Lemma 2 to the Gaussian process $\mathcal{G}$, we first need this approximation lemma.

Lemma A.1.

$$\sup_{\delta} \left| \sum_{j=3}^{J} \frac{\delta^{j-1}}{\sqrt{j!}} \left(\zeta_j + \frac{h_* \delta_*^{j-1}}{\sqrt{j!}} \right) - \sum_{j=3}^{\infty} \frac{\delta^{j-1}}{\sqrt{j!}} \left(\zeta_j + \frac{h_* \, \delta_*^{j-1}}{\sqrt{j!}} \right) \right| \xrightarrow{\mathbb{P}} 0.$$

Proof This requires a bound on the tail of the sum

$$\sum_{j=J+1}^{\infty} \frac{\delta^{j-1}}{\sqrt{j!}} \left(\zeta_j + \frac{h_* \, \delta_*^{j-1}}{\sqrt{j!}} \right) = \sum_{j>J} \left(\frac{\delta^{j-1}}{\sqrt{j!}} \zeta_j \right) + h_* \sum_{j>J} \frac{(\delta \delta_*)^{j-1}}{j!}.$$

By triangle inequality, we can bound the two pieces of this error separately. The first part is

$$\sup_{\delta} \left| \sum_{j>J} \left(\frac{\delta^{j-1}}{\sqrt{j!}} \zeta_j \right) \right| \le \sum_{j>J} \frac{a^{j-1}}{\sqrt{j!}} |\zeta_j| \xrightarrow{\mathbb{P}} 0$$

because the expectation is the tail of a convergent series.

The second part of the bound is

$$\sup_{\delta} \left| h_* \sum_{j>J} \frac{(\delta \delta_*)^{j-1}}{j!} \right| \le \left| \frac{h_*}{J} \right| \sum_{j>J} \frac{(a \delta_*)^{j-1}}{(j-1)!} \to 0.$$

Thus proving Lemma A.1.

As in Billingsley (1999) Theorem 3.2, the convergence of the average of the Z_δ follows because

$$\sqrt{n}m_1 = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_\delta(\widehat{v}_t) \xrightarrow{\mathbb{P}} \sum_{j=3}^{J}\left(\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_k(\widehat{v}_t)\right]\frac{\delta^{j-1}}{j!}\right)$$

by Lemma 2, and then

$$\sum_{j=3}^{J}\left(\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_k(\widehat{v}_t)\right]\frac{\delta^{j-1}}{j!}\right) \rightsquigarrow \sum_{j=3}^{J}\frac{\delta^{j-1}}{\sqrt{j!}}\left(\zeta_j + \frac{h_*\delta_*^{j-1}}{\sqrt{j!}}\right)$$

by Lemma 3. In turn, this converges by Lemma A.1 to

$$\sum_{j=3}^{\infty}\frac{\delta^{j-1}}{\sqrt{j!}}\left(\zeta_j + \frac{h_*\,\delta_*^{j-1}}{\sqrt{j!}}\right).$$

From Lemma 1, the likelihood ratio as a function of δ is

$$q_n(\delta) = \left[\frac{\sqrt{n}m_1}{\sqrt{m_2 - 1 - \delta^2 m_c^2/2}}\right]^2 \{\delta m_1 > 0\} + o_P(1)$$

The denominator of this fraction will converge in probability.

$$m_2 = \frac{1}{n}\sum_{t=1}^{n} Z_\delta(\widehat{v}_t)^2 \xrightarrow{\mathbb{P}} \mathbb{E}Z_\delta(u_t)^2 = \delta^{-2}\left(\mathsf{e}^{\delta^2} - 1\right) \tag{15}$$

where we use that $\widehat{v}_t - u_t = O_P(n^{-1/2})$. Likewise,

$$m_c = \frac{1}{n}\sum_{t=1}^{n} \widehat{v}_t Z_\delta(\widehat{v}_t) \xrightarrow{\mathbb{P}} \mathbb{E}u_t Z_\delta(u_t) = 1. \tag{16}$$

Therefore, by the continuous mapping theorem and our limit distribution for $\sqrt{n}m_1$

$$\frac{\sqrt{n}m_1}{\sqrt{m_2 - 1 - \delta^2 m_c^2/2}} \rightsquigarrow \mathcal{G}(\delta) = \sum_{k=3}^{\infty}\frac{\delta^{k-1}}{\sqrt{k!}}\left[\zeta_k + \frac{h^*\delta^{*\,k-1}}{\sqrt{k!}}\right]\left[\frac{1}{\delta^2}\left(\mathsf{e}^{\delta^2} - 1 - \delta^2 - \frac{\delta^4}{2}\right)\right]^{-1/2}.$$

The Likelihood ratio statistic is then the squared value of this Gaussian process when it has the same sign as δ,

$$q_\delta \rightsquigarrow \begin{cases} [\mathcal{G}_+]^2 & \text{for } \delta > 0, \\ [\mathcal{G}_-]^2 & \text{for } \delta < 0. \end{cases}$$

The assumption in the preceding argument is that π is in a shrinking neighborhood of 0. However, the mixture distribution has identifiability issues that mean that a neighborhood of $\delta = 0$ and $\pi > 0$ is still close to the null hypothesis of a single normal distribution.

This case is explored in great detail in Cho and White (2007), and we will not reproduce their arguments here. Most cases are handled by taking the limit of the Gaussian process as $\delta \to 0$.

$$\mathcal{G}(\delta) = \frac{\delta^2}{\sqrt{6}}\left[\zeta_3 + \frac{h_*\,\delta_*^2}{\sqrt{6}}\right]\left[\frac{1}{\delta^2}\left(\mathrm{e}^{\delta^2} - 1 - \delta^2 - \frac{\delta^4}{2}\right)\right]^{-1/2} + O_P(\delta) \tag{17}$$

$$= \left[\zeta_3 + \frac{h_*\,\delta_*^2}{\sqrt{6}}\right]\left(\frac{6}{\delta^6}\left[\frac{\delta^6}{6} + O(\delta^8)\right]\right)^{-1/2} + O_P(\delta) \tag{18}$$

$$= \zeta_3 + \frac{h_*\,\delta_*^2}{\sqrt{6}} + O_P(1). \tag{19}$$

This limit is the same as we approach 0 from the left or right so that if $\zeta_3 + \frac{h_*\,\delta_*^2}{\sqrt{6}}$ is positive or negative, there is a sequence of positive or negative δ's respectively converging to the limit. This implies

$$\limsup_{\delta\to 0} q_\delta = \left(\zeta_3 + \frac{h_*\,\delta_*^2}{\sqrt{6}}\right)^2 .$$

Notice that by definition $q_0 = 0$ and thus does not converge to $\mathcal{G}(0)$. If we are looking at the behavior for δ in a small neighborhood of zero then the identifiability issues come into play and we have a Fisher information of 0. There is almost always a local maximum near $\delta = 0$ with $\pi > 0$ where the likelihood is approximately $\frac{1}{\sqrt{n}}\sum_{t=1}^n H_3(\widehat{v}_t)/\sqrt{6}$. Therefore, the supremum of $\mathcal{G}_+(\delta)$ over all δ is always going to be larger than this local maximum.

However, there is one additional exception: when $\pi = 1/2$. In this neighborhood, the skewness of the distribution is nearly zero and so a local maximum occurs for $\delta = O(n^{-1/8})$. As in Cho and White, the leading term is actually $G = \zeta_4 + \sqrt{6}\ h_*\delta_*^3/12$. Therefore, our asymptotic likelihood process includes an additional term which comes from the excess kurtosis of the sample.

A.4.1. In a neighborhood of $\delta = 0$

Setting up the likelihoods as in Lemma 1 and then taking the terms for small δ and a fixed $\pi > 0$, the MLE estimates of the normal parameters are $\hat{\mu} = \bar{y} - \sigma\pi\delta$ and $\hat{\sigma}^2 = s^2/(1 + \pi(1-\pi)\delta^2)$. As a convenience set $\gamma = s^2/\hat{\sigma}^2 - 1$.

$$\begin{aligned} q_\pi(\delta) &= n\log(1+\gamma) - n\gamma + \\ &\quad + 2\sum_{t=1}^n \log\left[1 + \gamma\sum_{k=0}^\infty \frac{\delta^k}{(k+2)!} H_{k+2}\left(\widehat{v}_t\sqrt{1+\gamma}\right)\left((1-\pi)^{k+1} - (-\pi)^{k+1}\right)\right] \\ &= -\frac{n\gamma^2}{2} + \frac{n\gamma^3}{3} + 2\sum_{t=1}^n \log\left[1 + \frac{\gamma}{2}H_2\left(\widehat{v}_t\sqrt{1+\gamma}\right) + \right. \\ &\quad \left. + \frac{\delta(1-2\pi)\gamma}{6}H_3\left(\widehat{v}_t\sqrt{1+\gamma}\right) + \frac{\delta^2(1-3\pi+3\pi^2)\gamma}{24}H_4\left(\widehat{v}_t\sqrt{1+\gamma}\right) + \cdots\right] \end{aligned}$$

This expansion illustrates the behavior for small values of δ. The second Hermite polynomial terms are cancelled because $\sum_t H_2(\widehat{v}_t) = 0$. Thus the leading term in this expansion is

$$\frac{\delta(1-2\pi)\gamma}{6} \sum_{t=1}^{n} H_3\left(\widehat{v}_t\right)$$

which is $O_P(1)$ when $\delta = O(n^{-1/6})$. However, if $\pi = 1/2$, then even this term is zero in the expansion. Thus the leading term becomes.

$$\frac{\delta^2(1-3\pi+3\pi^2)\gamma}{24} \sum_{t=1}^{n} H_4(\widehat{v}_t)$$

which is $O_P(1)$ when $\delta = O(n^{-1/8})$.

Appendix B: Hermite Polynomials

We define the Hermite polynomial $H_j\left(x\right)$ in terms of the generating function

$$e^{x\delta-\delta^2/2} = \sum_{j=0}^{\infty} \frac{H_j\left(x\right)}{j!}\delta^j. \tag{20}$$

Specifically, the first 6 Hermite polynomials are

$$\begin{aligned} H_0\left(x\right) &= 1 & H_3\left(x\right) &= x^3 - 3x \\ H_1\left(x\right) &= x & H_4\left(x\right) &= x^4 - 6x^2 + 3 \\ H_2\left(x\right) &= x^2 - 1 & H_5\left(x\right) &= x^5 - 10x^3 + 15x \end{aligned}$$

The Hermite polynomials are an orthogonal series with respect to the standard normal density. In particular, for a standard normal Z,

$$\mathbb{E}\left[H_j\left(Z\right)H_k\left(Z\right)\right] = \begin{cases} j! & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases} \tag{21}$$

This can be established via the arguments described in Lebedev (1965) pages 60–76 using a generating series.

For the mixture of normals, we have similar moments.

Lemma B.1. *Assuming u_t is a standard normal and $s_t\delta$ is either 0 or δ with expectation $h/\sqrt{n}$, then for $k > 2$*

$$\mathbb{E}H_k(u_t + s_t\delta) = \frac{h\delta^{k-1}}{\sqrt{n}} \tag{22}$$

$$\mathbb{E}(u_t + s_t\delta)H_k(u_t + s_t\delta) = \frac{h}{\sqrt{n}}\left(\delta^k + k\delta^{k-2}\right) \tag{23}$$

$$\operatorname{Var} H_k(u_t + s_t\delta) = k! + O(n^{-1/2}) \tag{24}$$

$$\operatorname{Cov}\left(H_j(u_t + s_t\delta),\ H_k(u_t + s_t\delta)\right) = O(n^{-1/2}) \tag{25}$$

for $j < k$.

Proof of Lemma B.1 The $H_k(x)$ form a basis that is orthogonal with respect to the standard normal measure. The shifted measure has density $\exp\left[x\delta - \delta^2\right]$. This means that

$$\begin{aligned}\mathbb{E}H_k(u_t + \delta) &= \int H_k(x)\exp\left[x\delta - \delta^2\right]\phi(x)\,dx \\ &= \sum_{j=0}^{\infty}\frac{\delta^j}{j!}\int H_k(x)H_j(x)\phi(x)\,dx = \delta^k\end{aligned}$$

which proves (22).

The equation (23) follows from the standard identity $xH_k(x) = H_{k+1}(x) + kH_{k-1}(x)$ and (22).

The covariances are

$$\begin{aligned}\operatorname{Cov}(H_j(u_t{+}s_t\delta), H_k(u_t{+}s_t\delta)) &= \mathbb{E}\left[\operatorname{Cov}(H_j(u_t + s_t\delta), H_k(u_t + s_t\delta)) \mid s_t\right] + \delta^{j+k}\operatorname{Var}(s_t) \\ &= \operatorname{Cov}(H_j(u_t), H_k(u_t)) + \\ +\mathbb{E}(s_t)\left(\operatorname{Cov}(H_j(u_t + \delta), H_k(u_t + \delta)) - \operatorname{Cov}(H_j(u_t), H_k(u_t))\right) &+ \delta^{j+k-1}hn^{-1/2}(\delta - hn^{-1/2}) \\ &= \operatorname{Cov}(H_j(u_t), H_k(u_t)) + O(n^{-1/2})\end{aligned}$$

which proves (24) and (25).

Appendix C: Monte Carlo Sampling Error

To construct the Monte Carlo sampling error, we divide our 1000 simulations into 10 blocks of 100 simulations each. For each block, compute the average likelihood ratio $Q_{n,b}$, so the reported value for Q_n is $\overline{Q}_n = \frac{1}{10}\sum_{b=1}^{10} Q_{n,b}$. The variance of $Q_{n,b}$ is then $s^2_{Q,b} = \frac{1}{10}\sum_{b=1}^{10}\left(Q_{n,b} - \overline{Q}_n\right)^2$ and the reported Monte Carlo sampling error is the standard error of $\overline{Q}_n$, which equals $\frac{s_{Q,b}}{\sqrt{10}}$.

Acknowledgments

The authors thank Jim Hamilton, Dick Startz and members of the Santa Barbara Econometrics Research Group for helpful comments and Ben Hansen for invaluable research assistance.

References

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 2nd ed. A Wiley-Interscience Publication.

Carter, A. and Steigerwald, D. (2013). Markov regime-switching tests: Asymptotic critical values. *Journal of Econometric Methods* **2** 25–34.

Carter, A. and Steigerwald, D. (2018). "Asymptotic power of a likelihood ratio test for a mixture of normal distributions": Supplemental calculations. http://www.pstat.ucsb.edu/faculty/carter/publications.htm.

Chen, H., Chen, J. and Kalbfleisch, J. D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **63** 19–29.

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

Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary arma processes. *Ann. Statist.* **27** 1178–1209.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43.

Gassiat, E. (2002). Likelihood ratio inequalities with applications to various mixtures. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 897–906. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

Hall, P. and Stewart, M. (2005). Theoretical analysis of power in a two-component normal mixture model. *J. Statist. Plann. Inference* **134** 158–179.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 413–430.

Hartigan, J. (1985). A failure of likelihood ratio asymptotics for normal mixtures. In L. M. Le Cam and R. A. Olshen, eds., *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*.

Kasahara, H., Okimoto, T. and Shimotsu, K. (2014). Modified quasi-likelihood ratio test for regime switching. *Jpn. Econ. Rev.* **65** 25–41.

Lebedev, N. N. (1965). *Special Functions and their Applications*. Revised English edition. Translated and edited by Richard A. Silverman. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Statist.* **31** 807–832.

Liu, X. and Shao, Y. (2004). Asymptotics for the likelihood ratio test in a two-component normal mixture model. *J. Statist. Plann. Inference* **123** 61–81.