Asymptotic Power of a Likelihood Ratio Test for a Mixture of Normal Distributions: Supplemental Calculations

Andrew V. Carter
University of California, Santa Barbara

Douglas G. Steigerwald
University of California, Santa Barbara

December 13, 2018

1 The likelihood for small δ

The likelihood is

$$L(\beta,\sigma,\delta,\pi) = \sum_{t=1}^{n} \log\left[\frac{1-\pi}{\sigma}\phi\left(v_t\right) + \frac{\pi}{\sigma}\phi\left(v_t-\delta\right)\right] \tag{1}$$

where $v_t = \frac{y_t - x_t^{\mathrm{T}}\beta}{\sigma}$. The test statistic subtracts off the likelihood ratio under the null which produces

$$q_\pi(\beta,\sigma,\delta) = n\log\frac{s^2}{\sigma^2} + 2\sum_{t=1}^{n}\log\left[(1-\pi)\frac{\phi\left(v_t\right)}{\phi(\widehat{v}_t)} + \pi\frac{\phi\left(v_t-\delta\right)}{\phi(\widehat{v}_t)}\right] \tag{2}$$

The necessary algebra here is that

$$v_t^2 = \frac{(y_t - x_t b + x_t(b-\beta))^2}{\sigma^2} = \frac{s^2}{\sigma^2}\left(\widehat{v}_t + \frac{x_t(b-\beta)}{s}\right)^2$$

$$\log\frac{\phi\left(v_t\right)}{\phi(\widehat{v}_t)} = \frac{\widehat{v}_t^2}{2}\left(1-\frac{s^2}{\sigma^2}\right) + \left(\frac{x_t(\beta-b)}{\sigma}\right)\frac{s\widehat{v}_t}{\sigma} - \frac{\left[x_t(\beta-b)\right]^2}{2\sigma^2}$$

$$\log\frac{\phi\left(v_t-\delta\right)}{\phi(\widehat{v}_t)} = \frac{\widehat{v}_t^2}{2}\left(1-\frac{s^2}{\sigma^2}\right) + \left(\frac{x_t(\beta-b)}{\sigma}+\delta\right)\frac{s\widehat{v}_t}{\sigma} - \frac{1}{2}\left[\frac{x_t(\beta-b)}{\sigma}+\delta\right]^2$$

Implying that the likelihood is

$$
\begin{aligned}
q_\pi(\beta,\sigma,\delta) &= n\log\frac{s^2}{\sigma^2}+\sum_{t=1}^{n}\frac{\widehat{v}_t^2}{2}\left(1-\frac{s^2}{\sigma^2}\right)+\\
&+2\sum_{t=1}^{n}\log\left[(1-\pi)\exp\left[-\left(\frac{x_t(b-\beta)}{\sigma}\right)\frac{s\widehat{v}_t}{\sigma}-\frac{[x_t(b-\beta)]^2}{2\sigma^2}\right]+\right.\\
&\left.+\pi\exp\left(-\left(\frac{x_t(b-\beta)}{\sigma}-\delta\right)\frac{s\widehat{v}_t}{\sigma}-\frac{1}{2}\left[\frac{x_t(b-\beta)}{\sigma}-\delta\right]^2\right)\right] \qquad (3)
\end{aligned}
$$

We will use our Hermite polynomial expansion on these exponential terms

$$
\exp\left[\left(\frac{x_t(\beta-b)}{\sigma}\right)\frac{s\widehat{v}_t}{\sigma}-\frac{[x_t(\beta-b)]^2}{2\sigma^2}\right]=\sum_{k=0}^{\infty}H_k\left(\frac{\widehat{v}_t\,s}{\sigma}\right)\frac{\tau_t^k}{k!} \qquad (4)
$$

$$
\exp\left(\left(\frac{x_t(\beta-b)}{\sigma}+\delta\right)\frac{s\widehat{v}_t}{\sigma}-\frac{1}{2}\left[\frac{x_t(\beta-b)}{\sigma}+\delta\right]^2\right)=\sum_{k=0}^{\infty}H_k\left(\frac{\widehat{v}_t\,s}{\sigma}\right)\frac{(\tau_t+\delta)^k}{k!} \qquad (5)
$$

where we define $\tau_t = [x_t(\beta-b)]/\sigma$.

$$
q_\pi(\beta,\sigma,\delta)=n\log\frac{s^2}{\sigma^2}+n\left(1-\frac{s^2}{\sigma^2}\right)+2\sum_{t=1}^{n}\log\left[\sum_{k=0}^{\infty}\frac{1}{k!}H_k\left(\frac{\widehat{v}_t\,s}{\sigma}\right)\left((1-\pi)\tau_t^k+\pi(\tau_t+\delta)^k\right)\right]. \qquad (6)
$$

Presumably, τ_t and δ are small here so we can approximate this by the first three terms in the sum.

$$
\sum_{k=0}^{\infty}\frac{1}{k!}H_k\left(\frac{\widehat{v}_t\,s}{\sigma}\right)\left((1-\pi)\tau_t^k+\pi(\tau_t+\delta)^k\right)=1+\frac{\widehat{v}_t\,s}{\sigma}(\tau_t+\pi\delta)+\frac{1}{2}\left(\frac{\widehat{v}_t^2\,s^2}{\sigma^2}-1\right)\left(\tau_t^2+2\pi\tau_t\delta+\pi\delta^2\right)+\cdots
$$

1.1 Maximizing over β

Maximizing (6) over all values of β is nearly the same as maximizing

$$
\begin{aligned}
&\sum_{t=1}^{n}\log\left[1+\frac{\widehat{v}_t\,s}{\sigma}(\tau_t+\pi\delta)+\frac{1}{2}\left(\frac{\widehat{v}_t^2\,s^2}{\sigma^2}-1\right)\left(\tau_t^2+2\pi\tau_t\delta+\pi\delta^2\right)\right]\\
&=\sum_{t=1}^{n}\log\left[1+\frac{\widehat{v}_t\,s}{\sigma}(\tau_t+\pi\delta)+\frac{1}{2}\left(\frac{\widehat{v}_t^2\,s^2}{\sigma^2}-1\right)\left([\tau_t+\delta\pi]^2+\pi(1-\pi)\delta^2\right)\right]
\end{aligned}
$$

If we approximate $\log(1+a) \approx a$, then the largest term

$$\frac{s}{\sigma}\sum_{t=1}^{n} \widehat{v}_t(\tau_t + \pi\delta) = 0$$

because $\widehat{v}_t$ is orthogonal to column space of x_t which includes the τ_t. Therefore, we need to include the quadratic term,

$$\sum_{t=1}^{n} \log\left[1 + \frac{\widehat{v}_t\, s}{\sigma}(\tau_t + \pi\delta) + \frac{1}{2}\left(\frac{\widehat{v}_t^2\, s^2}{\sigma^2} - 1\right)\left(\tau_t^2 + 2\pi\tau_t\delta + \pi\delta^2\right)\right]$$

$$= \frac{1}{2}\sum_{t=1}^{n}\left(\frac{\widehat{v}_t^2\, s^2}{\sigma^2} - 1\right)\left([\tau_t + \delta\pi]^2 + \pi(1-\pi)\delta^2\right) - \frac{s^2}{2\sigma^2}\sum_{t=1}^{n}\widehat{v}_t^2(\tau_t + \pi\delta)^2 + o(\delta^2)$$

$$= -\frac{1}{2}\sum_{t=1}^{n}(\tau_t + \pi\delta)^2 + \frac{\delta^2\pi(1-\pi)}{2}\sum_{t=1}^{n} H_2\left(\widehat{v}_t\frac{s}{\sigma}\right) + o(\delta^2)$$

This is clearly maximized at $\tau_t = -\pi\delta$ which is conveniently within the model space so that there exists a $\hat{\beta}$ such that

$$x_t\hat{\beta} = x_t b - \sigma\pi\delta.$$

1.2 Maximizing over σ

Taking $\tau_t = -\pi\delta$ in (6), the likelihood ratio is

$$q_\pi(\sigma,\delta) = n\log\frac{s^2}{\sigma^2} + n\left(1 - \frac{s^2}{\sigma^2}\right) +$$
$$+\, 2\sum_{t=1}^{n}\log\left[1 + \sum_{k=2}^{\infty}\frac{\pi(1-\pi)\delta^k}{k!}H_k\left(\widehat{v}_t\frac{s}{\sigma}\right)\left((1-\pi)^{k-1} - [-\pi]^{k-1}\right)\right]. \quad (7)$$

For simplicity, we will denote $\gamma = s^2/\sigma^2 - 1$, and use Lemma 2.1

$$H_2\left(\widehat{v}_t\frac{s}{\sigma}\right) = H_2(\widehat{v}_t)(1+\gamma) + \gamma$$
$$H_3\left(\widehat{v}_t\frac{s}{\sigma}\right) = H_3(\widehat{v}_t)(1 + \frac{3}{2}\gamma) + 3\gamma\,\widehat{v}_t + O(\gamma^2),$$

Thus, the first terms in this expansion are

$$q_\pi(\sigma,\delta) = n\log(1+\gamma) - n\gamma +$$
$$+\, 2\sum_{t=1}^{n}\log\left[1 + \frac{\pi(1-\pi)\delta^2}{2}\left(H_2(\widehat{v}_t)(1+\gamma) + \gamma\right) + \cdots\right]$$

Taking the derivative we get

$$\frac{\partial}{\partial\sigma}q_\pi(\sigma,\delta) = -\frac{n\gamma}{1+\gamma}+\pi(1-\pi)\delta^2\sum_{t=1}^{n}\frac{H_2(\widehat{v}_t)+1+\cdots}{1+\frac{\pi(1-\pi)\delta^2}{2}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)+\cdots}$$

$$= -\frac{n\gamma}{1+\gamma}+n\pi(1-\pi)\delta^2-\frac{\pi^2(1-\pi)^2\delta^4}{2}\sum_{t=1}^{n}\big(H_2(\widehat{v}_t)^2(1+\gamma)+H_2(\widehat{v}_t)(1+2\gamma)+\gamma\big)+O(n\delta^6)$$

$$= -\frac{n\gamma}{1+\gamma}+n\pi(1-\pi)\delta^2-n(1+\gamma)\pi^2(1-\pi^2)\delta^4\left[\frac{1}{2n}\sum_{t=1}^{n}H_2(\widehat{v}_t)^2\right]-\frac{n\gamma\pi^2(1-\pi)^2\delta^4}{2}+O(n\delta^6)$$

where we use $\sum_t H_2(\widehat{v}_t) = 0$. This derivative is nearly zero when $\gamma = \pi(1-\pi)\delta^2$. Therefore,

$$q_\pi(\hat{\sigma},\delta) = n\log(1+\gamma)-n\gamma+2\sum_{t=1}^{n}\log\left[1+\frac{\gamma}{2}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)+\cdots\right]$$

$$= -\frac{n\gamma^2}{2}+\sum_{t=1}^{n}\big(H_2(\widehat{v}_t)(\gamma+\gamma^2)+\gamma^2\big)-\frac{\gamma^2}{4}\sum_{t=1}^{n}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)^2+O(n\delta^6)$$

$$= -\frac{n\gamma^2}{2}+n\gamma^2-\frac{n\gamma^2(1+\gamma)^2}{2}\left[\frac{1}{2n}\sum_{t=1}^{n}H_2(\widehat{v}_t)^2\right]+O(n\delta^6)$$

$$= \frac{n\pi^2(1-\pi)^2\delta^4}{2}\left[1-\frac{1}{2n}\sum_{t=1}^{n}H_2(\widehat{v}_t)^2\right]+O(n\delta^6) \quad (8)$$

However, this last term is actually small because $\mathbb{E}H_2(\widehat{v}_t)^2 = 2$ so that the leading term in the likelihood is $O_P(n\delta^6+\sqrt{n}\delta^4)$.

1.2.1 Expansion in γ

We would like to expand the likelihood expression from (8) out as a Taylor expansion around $\gamma = 0$.

$$n\log(1+\gamma)-n\gamma+2\sum_{t=1}^{n}\log\left[1+\frac{\gamma}{2}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)\right]$$

$$= -\frac{n\gamma^2}{2}+\frac{n\gamma^3}{3}-\frac{n\gamma^4}{4}+\sum_{t=1}^{n}\Bigg[\gamma\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)-\frac{\gamma^2}{4}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)^2+$$

$$+\frac{\gamma^3}{12}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)^3-\frac{\gamma^4}{32}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)^4\Bigg]+O_P(n\gamma^5) \quad (9)$$

The terms in this series for the powers of the second Hermite polynomial can be simplified because $\sum H_2(\widehat{v}_t) = 0$

$$\sum_{t=1}^{n} \gamma \left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right) = n\gamma^2$$

$$\begin{aligned}\frac{\gamma^2}{4}\sum_{t=1}^{n}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)^2 &= \frac{\gamma^2}{4}\sum_{t=1}^{n}\left(H_2(\widehat{v}_t)^2(1+\gamma)^2+2\gamma(1+\gamma)H_2(\widehat{v}_t)+\gamma^2\right)\\ &= \frac{\gamma^2+2\gamma^3+\gamma^4}{4}\left[\sum_{t=1}^{n} H_2(\widehat{v}_t)^2\right]+\frac{n\gamma^4}{4}\end{aligned}$$

Then

$$\frac{\gamma^3}{12}\sum_{t=1}^{n}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)^3 = \frac{\gamma^3}{12}\left[\sum_{t} H_2(\widehat{v}_t)^3\right]+\frac{\gamma^4}{4}\sum_{t=1}^{n}\left[H_2(\widehat{v}_t)^3+H_2(\widehat{v}_t)^2\right]+O_P(n\gamma^5)$$

$$\frac{\gamma^4}{32}\sum_{t=1}^{n}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)^4 = \frac{\gamma^4}{32}\sum_{t=1}^{n}\left[H_2(\widehat{v}_t)^4\right]+O_P(n\gamma^5)$$

Thus,

$$\begin{aligned}&n\log(1+\gamma)-n\gamma+2\sum_{t=1}^{n}\log\left[1+\frac{\gamma}{2}\left(H_2(\widehat{v}_t)(1+\gamma)+\gamma\right)\right]\\ &\quad= -\frac{n\gamma^2}{2}+n\gamma^2-\frac{\gamma^2+2\gamma^3+\gamma^4}{4}\left[\sum_{t=1}^{n} H_2(\widehat{v}_t)^2\right]-\frac{n\gamma^4}{4}+\\ &\quad\quad+\frac{n\gamma^3}{3}+\frac{\gamma^3}{12}\left[\sum_{t} H_2(\widehat{v}_t)^3\right]+\frac{\gamma^4}{4}\sum_{t=1}^{n}\left[H_2(\widehat{v}_t)^3+H_2(\widehat{v}_t)^2\right]+\\ &\quad\quad-\frac{n\gamma^4}{4}-\frac{\gamma^4}{32}\sum_{t=1}^{n}\left[H_2(\widehat{v}_t)^4\right]+O_P(n\gamma^5)\\ &= \frac{n\gamma^2}{2}\left(1-\frac{1}{2n}\left[\sum_{t=1}^{n} H_2(\widehat{v}_t)^2\right]\right)+\frac{n\gamma^3}{3}\left(1-\frac{3}{2n}\sum_{t=1}^{n} H_2(\widehat{v}_t)^2+\frac{1}{4n}\sum_{t=1}^{n} H_2(\widehat{v}_t)^3\right)+\\ &\quad-\frac{n\gamma^4}{2}\left(1-\frac{1}{2n}\sum_{t=1}^{n} H_2(\widehat{v}_t)^3+\frac{1}{16n}\sum_{t=1}^{n} H_2(\widehat{v}_t)^4\right)+O_P(n\gamma^5) \quad (10)\end{aligned}$$

These powers of the polynomial terms can be written as equivalent linear combinations of higher order Hermite polynomials. In particular, a bit of algebra shows that

$$\begin{aligned}
H_2(x)^2 &= H_4(x) + 4H_2(x) + 2 \\
H_2(x)^3 &= H_6(x) + 12H_4(x) + 30H_2(x) + 8 \\
H_2(x)^4 &= H_8(x) + 24H_6(x) + 156H_4(x) + 272H_2(x) + 60
\end{aligned}$$

Then we will define

$$\xi_k = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} H_k(\widehat{v}_t)$$

where for large n, $\xi_k = O_P(1)$. This leads to the expressions

$$\begin{aligned}
\sum_{t=1}^{n} H_2(\widehat{v}_t)^2 &= \sqrt{n}\xi_4 + 2n \\
\sum_{t=1}^{n} H_2(\widehat{v}_t)^3 &= 8n + O_P(\sqrt{n}) \\
\sum_{t=1}^{n} H_2(\widehat{v}_t)^4 &= 60n + O_P(\sqrt{n})
\end{aligned}$$

Plugging these in

$$\begin{aligned}
n\log(1+\gamma) - n\gamma + 2\sum_{t=1}^{n} \log\left[1 + \frac{\gamma}{2}\left(H_2(\widehat{v}_t)(1+\gamma) + \gamma\right)\right] \\
= -\frac{\sqrt{n}\gamma^2}{4}\xi_4 - \frac{3n\gamma^4}{8} + O_P(n\gamma^5) + O_P(\sqrt{n}\gamma^3). \quad (11)
\end{aligned}$$

1.3 Maximizing over δ

Using a slight abuse of our notation, we will continue to use $\gamma = \pi(1-\pi)\delta^2$. Then the likelihood ratio function is

$$q_\pi(\hat{\beta}, \hat{\sigma}, \delta) = n\log(1+\gamma) - n\gamma + 2\sum_{t=1}^{n} \log\left[1 + \frac{\gamma}{2} H_2(\widehat{v}_t s/\hat{\sigma}) + \frac{(1-2\pi)\gamma\delta}{6} H_3(\widehat{v}_t s/\hat{\sigma}) + O(\delta^4)\right]$$

In section 1.2.1, we have an approximation for the terms that include the second Hermite polynomial so

$$q_\pi(\hat{\beta}, \hat{\sigma}, \delta) = O_P(n\delta^6) + O_P(\sqrt{n}\delta^4) + \frac{(1-2\pi)\gamma\delta}{3} \sum_{t=1}^{n} H_3(\widehat{v}_t s/\hat{\sigma}) - \frac{(1-2\pi)^2\gamma^2\delta^2}{36} H_3(\widehat{v}_t s/\hat{\sigma})$$

For the third Hermite polynomial

$$\sum_{t=1}^{n} H_3(\widehat{v}_t s/\sigma) = \sqrt{n}\left(1+\frac{3\gamma}{2}\right)\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_3(\widehat{v}_t)\right] + O_P(\sqrt{n}\gamma^2)$$

because the $\widehat{v}_t$ sum to 0. This implies that the contribution from these terms are

$$\sum_{t=1}^{n}\left(\frac{(1-2\pi)\gamma\delta}{3}H_3(\widehat{v}_t s/\hat{\sigma}) - \frac{(1-2\pi)^2\gamma^2\delta^2}{36}H_3(\widehat{v}_t s/\hat{\sigma})^2\right)$$
$$= \frac{\sqrt{n}(1-2\pi)\gamma\delta}{3}\xi_3 - \frac{n(1-2\pi)^2}{6}\gamma^2\delta^2 + O_P(n\delta^8 + \sqrt{n}\delta^5) \quad (12)$$

There is a term that comes from the product which is negligible

$$\sum_{t=1}^{n}\frac{(1-2\pi)\gamma^2\delta}{9}H_2(\widehat{v}_t s/\hat{\sigma})H_3(\widehat{v}_t s/\hat{\sigma}) = O_P(\sqrt{n}\delta^5)$$

Therefore,

$$q_\pi(\hat{\beta},\hat{\sigma},\delta) = \frac{\sqrt{n}(1-2\pi)\gamma\delta}{3}\xi_3 - \frac{n(1-2\pi)^2}{6}\gamma^2\delta^2 + O_P(n\delta^8 + \sqrt{n}\delta^4) \quad (13)$$

Completing the square we can see that this is

$$q_\pi(\hat{\beta},\hat{\sigma},\delta) = -\frac{1}{6}\left(\sqrt{n}(1-2\pi)\gamma\delta - \xi_3\right)^2 + \frac{\xi_3^2}{6} + O_P(n\delta^8 + \sqrt{n}\delta^4) \quad (14)$$

which has a maximum over δ of

$$q_\pi(\hat{\beta},\hat{\sigma},\hat{\delta}) = \frac{\xi_3^2}{6} + o_P(n^{-1/3}).$$

It also shows that the maximizing value of δ is

$$\sqrt{n}(1-2\pi)\gamma\hat{\delta} = \xi_3$$
$$\Longrightarrow \hat{\delta}^3 = n^{-1/2}\left(\frac{\xi_3}{\pi(1-\pi)(1-2\pi)}\right)$$

which confirms that $\hat{\delta} = O_P(n^{-1/6})$.

Notice that the δ which maximizes the likelihood depends on which π we used but the maximum does not depend on the probability (with one exception described below). The likelihood surface has a ridge along which the gradient is 0. If we take this fixed π close to 0, then we can still get this same maximizing value. This heuristically explains why we take the continuous extension of the $G(\delta)$ process for $\delta = 0$.

1.4 Symmetric Case

The previous analysis breaks down if the probability π is fixed at $1/2$ because then $1-2\pi=0$ and the terms with the third Hermite polynomial drop out. In this case, we need to take a further expansion of our likelihood

$$q_\pi(\hat\beta,\hat\sigma,\delta) = n\log(1+\gamma) - n\gamma +$$
$$+ 2\sum_{t=1}^{n}\log\left[1+\frac{\gamma}{2}H_2(\widehat{v}_t s/\hat\sigma) + \frac{(1-3\pi+3\pi^2)\gamma\delta^2}{24}H_4(\widehat{v}_t s/\hat\sigma) + O(\delta^5)\right]$$

You can see that $\sum\gamma\delta^2 H_4(\widehat{v}_t s/\hat\sigma) = O_P(\sqrt{n}\delta^4)$ which will now be the size of the leading term in the expansion.

Calculations in Cho and White [2007] show that the likelihood is maximized at a δ in a $n^{-1/8}$ neighborhood of $\delta=0$ if the excess kurtosis term ξ_4 is positive. They showed that the likelihood ratio at this local maximum converges to $\xi_4^2/24$.

If $\xi_4<0$, then the maximum in this neighborhood of $\pi=1/2$ is at $\delta=0$, but there will be other local maximma that are larger.

2 Hermite Polynomial Bounds

The likelihood is a function of a re-scaled Hermite polynomial.

Lemma 2.1

$$H_1(\widehat{v}_t(1+\gamma)^{1/2}) = H_1(\widehat{v}_t)(1+\gamma/2) + O_P(\gamma^2) \tag{15}$$
$$H_2(\widehat{v}_t(1+\gamma)^{1/2}) = H_2(\widehat{v}_t)(1+\gamma) + \gamma \tag{16}$$
$$H_k(\widehat{v}_t(1+\gamma)^{1/2}) = H_k(\widehat{v}_t)\left(1+\frac{k\gamma}{2}\right) + \frac{k(k-1)\gamma}{2}H_{k-2}(\widehat{v}_t) + O(\gamma^2) \tag{17}$$

The results in (15) and (16) are immediate results of direct substitution. The interesting result follows from a Taylor expansion

$$H_k'(x) = kH_{k-1}(x)$$

so that

$$H_k(\widehat{v}_t(1+\gamma)^{1/2}) = H_k(\widehat{v}_t) + \frac{k\gamma\widehat{v}_t}{2}H_{k-1}(\widehat{v}_t) + O(\gamma^2).$$

Then the standard result

$$xH_k(x) = H_{k+1}(x) + kH_{k-1}(x)$$

applied to the second term implies

$$H_k(\widehat{v}_t(1+\gamma)^{1/2}) = H_k(\widehat{v}_t)\left(1+\frac{k\gamma}{2}\right) + \frac{\gamma k(k-1)}{2} H_{k-2}(\widehat{v}_t) + O(\gamma^2).$$

Lemma 2.2

$$H_2(x)^2 = H_4(x) + 4H_2(x) + 2 \tag{18}$$
$$H_2(x)^3 = H_6(x) + 12H_4(x) + 30H_2(x) + 8 \tag{19}$$
$$H_3(x)^2 = H_6(x) + 9H_4(x) + 18H_2(x) + 6 \tag{20}$$
$$H_2(x)H_3(x) = H_5(x) + 6H_3(x) + 6H_1(x) \tag{21}$$

This result follows via simple algebra on the definitions of the Hermite polynomials.

Combining these results from Lemma 2.2 and Lemma 2.1

$$\begin{aligned} H_2(\widehat{v}_t(1+\gamma)^{1/2})^2 &= H_4(\widehat{v}_t(1+\gamma)^{1/2}) + 4H_2(\widehat{v}_t(1+\gamma)^{1/2}) + 2 \\ &= H_4(\widehat{v}_t)(1+2\gamma) + 6\gamma H_2(\widehat{v}_t) + 4H_2(\widehat{v}_t)(1+\gamma) + 4\gamma + 2 + O(\gamma^2) \end{aligned} \tag{22}$$

For H_3,

$$\begin{aligned} H_3(\widehat{v}_t(1+\gamma)^{1/2})^2 &= H_6(\widehat{v}_t(1+\gamma)^{1/2}) + 9H_4(\widehat{v}_t(1+\gamma)^{1/2}) + 18H_2(\widehat{v}_t(1+\gamma)^{1/2}) + 6 \\ &= H_6(\widehat{v}_t)(1+3\gamma/2) + H_4(\widehat{v}_t)(9+33\gamma) + \\ &\quad + H_2(\widehat{v}_t)(18+72\gamma) + 18\gamma + 6 + O(\gamma^2) \end{aligned} \tag{23}$$

3 Approximating the Error using Residuals

The standardized residuals in the model are $\widehat{v}_t$, and presumably these are close to $w_t = u_t + \delta_* s_t$ which are the true residuals which have a mixture-normal distribution. This lemma says that they are close enough that the behaviour of the Hermite polynomials is the same for each. The key assumption is that the design has the typical asymptotics.

Condition 3.1 *The covariates x_t follow a distribution such that*

1. *The x_t are independent of the u_t and s_t.*

2. *The impact of the residuals on the least squares estimate are asymptotically negligible,*

$$\left[\mathbf{X}^\mathsf{T}\mathbf{X}\right]^{-1}\mathbf{X}^\mathsf{T}\vec{w} = O_P(n^{-1/2}).$$

These follow from standard conditions for asymptotic normality of regression estimators and the constraint that $\mathbb{E}\delta_* s_t = O(n^{-1/2})$.

Lemma 3.1 *Under Condition 3.1, as $n \to \infty$,*

$$\sum_{k=3}^{\infty} \frac{\delta^{k-1}}{k!} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(H_k(\widehat{v}_t) - H_k(w_t)\right) \right] \xrightarrow{\mathbb{P}} 0.$$

For $k \geq 3$, we have $H_k'(x) = kH_{k-1}(x)$ so that a Taylor expansion gives

$$\sum_{k=3}^{\infty} \frac{\delta^{k-1}}{k!} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(H_k(\widehat{v}_t) - H_k(w_t)\right) = \sum_{k=3}^{\infty} \frac{\delta^{k-1}}{n(k-1)!} \sum_{t=1}^{n} \sqrt{n}\left(\widehat{v}_t - w_t\right) H_{k-1}(w_t) + \\ + \sum_{k=3}^{\infty} \frac{\delta^{k-1}}{n(k-2)!} \sum_{t=1}^{n} \sqrt{n}\left(\widehat{v}_t - w_t\right)^2 H_{k-2}(v_t^*) \quad (24)$$

where v_t^* is between $\widehat{v}_t$ and w_t. We have shown that $\widehat{v}_t - w_t = O_P(n^{-1/2})$, but we will derive an explicit approximation of $\widehat{v}_t - w_t$.

The $\widehat{v}_t$ is a normalized residual from a projection onto the column space of the x_t (call this $\mathbf{P}_X$)

$$(\widehat{v}_t)_{t=1}^{n} = \vec{v} = \frac{(\mathbf{I} - \mathbf{P}_X)\,\vec{y}}{\sqrt{\frac{1}{n}\vec{y}^{\mathsf{T}}(\mathbf{I} - \mathbf{P}_X)\,\vec{y}}} = \frac{(\mathbf{I} - \mathbf{P}_X)(\vec{u} + \delta\vec{s})}{\sqrt{\frac{1}{n}(\vec{u} + \delta\vec{s})^{\mathsf{T}}(\mathbf{I} - \mathbf{P}_X)(\vec{u} + \delta\vec{s})}}$$

The estimate of the variance $\hat{\sigma}^2$ is (conditional on the s_t) a non-central χ^2 with $n-2$ degrees of freedom and $\delta^2\vec{s}^{\mathsf{T}}(\mathbf{I} - \mathbf{P}_X)\,\vec{s}$ as the non-centrality parameter. It is well known that this can be treated as a mixture of χ^2's with at least $n-2$ degrees of freedom. Thus, the marginal distribution (integrating over s_t) is also a mixture of χ^2's. The non-centrality parameter is small relative to $n-2$

$$\mathbb{E}\frac{1}{n}(\vec{u} + \delta\vec{s})^{\mathsf{T}}(\mathbf{I} - \mathbf{P}_X)(\vec{u} + \delta\vec{s}) \leq \frac{n-2}{n} + \frac{\delta^2}{n}\mathbb{E}M = 1 + \frac{h\delta}{\sqrt{n}} - \frac{2}{n}.$$

where M is the number of $s_t = 1$. The Central Limit Theorem then implies

$$\frac{1}{n}(\vec{u} + \delta\vec{s})^{\mathsf{T}}(\mathbf{I} - \mathbf{P}_X)(\vec{u} + \delta\vec{s}) - 1 = O_P(n^{-1/2}).$$

We can use this to approximate the denominator with a Taylor expansion $x^{-1/2} = 1 - (x-1)/2 + O((x-1)^2)$

$$\vec{v} = (\mathbf{I} - \mathbf{P}_X)(\vec{u} + \delta\vec{s}) \left[1 - \frac{1}{2}\left(\frac{1}{n}(\vec{u} + \delta\vec{s})^{\mathsf{T}}(\mathbf{I} - \mathbf{P}_X)(\vec{u} + \delta\vec{s}) - 1\right) + O_P(n^{-1})\right]$$

Implying that there are two important components to this approximation, an error in the estimate of σ and the error in the regression estimate. In particular,

$$\sqrt{n}(\vec{v}-\vec{u}-\delta\vec{s}) = \frac{\sqrt{n}}{2}\left(1-\frac{1}{n}\,(\vec{u}+\delta\vec{s})^{\mathsf{T}}\,(\mathbf{I}-\mathbf{P}_X)\,(\vec{u}+\delta\vec{s})\right)(\vec{u}+\delta\vec{s})+ \\ -\sqrt{n}\,\mathbf{P}_X\,(\vec{u}+\delta\vec{s})\left[1-\frac{1}{2}\left(\frac{1}{n}\,(\vec{u}+\delta\vec{s})^{\mathsf{T}}\,(\mathbf{I}-\mathbf{P}_X)\,(\vec{u}+\delta\vec{s})-1\right)\right]+O_P(n^{-1/2}) \tag{25}$$

This is our approximation of the differences between $\widehat{v}_t$ and w_t. To be concrete, we will define the projection of the w_t onto the column space of the covariates as having coefficient b_0 so that $\mathbf{P}_X\,(\vec{u}+\delta\vec{s})=\mathbf{X}b_0$.

We can plug this approximation into our first term of the Taylor expansion in (24),

$$\sum_{k=2}^{\infty}\frac{\delta^k}{k!}\frac{1}{n}\sum_{t=1}^{n}\sqrt{n}\,(\widehat{v}_t-w_t)\,H_k(w_t) = \\ \frac{\sqrt{n}}{2}\left(1-\frac{1}{n}\,(\vec{u}+\delta\vec{s})^{\mathsf{T}}\,(\mathbf{I}-\mathbf{P}_X)\,(\vec{u}+\delta\vec{s})\right)\sum_{k=2}^{\infty}\frac{\delta^k}{k!}\frac{1}{n}\sum_{t=1}^{n}w_tH_k(w_t)+ \\ -\sqrt{n}\left[1-\frac{1}{2}\left(\frac{1}{n}\,(\vec{u}+\delta\vec{s})^{\mathsf{T}}\,(\mathbf{I}-\mathbf{P}_X)\,(\vec{u}+\delta\vec{s})-1\right)\right]\sum_{k=2}^{\infty}\frac{\delta^k}{k!}\frac{1}{n}\sum_{t=1}^{n}x_t^{\mathsf{T}}b_0H_k\,(w_t) \tag{26}$$

The first factor is $\frac{\sqrt{n}}{2}\left(1-\frac{1}{n}\,(\vec{u}+\delta\vec{s})^{\mathsf{T}}\,(\mathbf{I}-\mathbf{P}_X)\,(\vec{u}+\delta\vec{s})\right)=O_P(1)$. The Law of Large Numbers implies that

$$\sum_{k=2}^{\infty}\frac{\delta^k}{k!}\frac{1}{n}\sum_{t=1}^{n}w_tH_k(w_t)\xrightarrow{\mathbb{P}}\sum_{k=2}^{\infty}\frac{\delta^k}{k!}\mathbb{E}(w_t)H_k(w_t), \tag{27}$$

and by Lemma B.1, $\mathbb{E}(w_t)H_k(w_t)=n^{-1/2}h_*\left(\delta_*^k+k\delta_*^{k-2}\right)$,and

$$\sum_{k=2}^{\infty}\frac{\delta^k}{k!}\mathbb{E}(w_t)H_k(w_t)=\frac{h}{\sqrt{n}}\left(\mathsf{e}^{\delta\delta_*}-1-\delta\delta_*+\delta/\delta_*\left[\mathsf{e}^{\delta\delta_*}-1\right]\right)\to 0.$$

Therefore, the first term in (26) is a product of these factors which also converges to 0 in probability.

In a similar fashion, the second term in (26) contains the factor

$$\left[1 - \frac{1}{2}\left(\frac{1}{n}\,(\vec{u} + \delta\vec{s})^{\mathsf{T}}\,(\mathbf{I} - \mathbf{P}_X)\,(\vec{u} + \delta\vec{s}) - 1\right)\right] = O_P(1).$$

The rest of the second factor is

$$\sqrt{n}\sum_{k=2}^{\infty}\frac{\delta^k}{k!}\frac{1}{n}\sum_{t=1}^{n} x_t^{\mathsf{T}} b_0 H_k\,(w_t) = \sqrt{n} b_0^{\mathsf{T}} \sum_{k=2}^{\infty}\frac{\delta^k}{k!}\frac{1}{n}\sum_{t=1}^{n} x_t^{\mathsf{T}} H_k\,(w_t)$$

Condition 3.1 implies that the b_0 vector are all $O_P(n^{-1/2})$ so the first vector has $O_P(1)$ coordinates, and the law of large numbers again

$$\frac{1}{n}\sum_{t=1}^{n} x_t H_k(w_t) \xrightarrow{\mathbb{P}} \mathbb{E}x_t \mathbb{E}H_k(w_t)$$

where by Lemma B.1

$$\sum_{k=2}^{\infty}\frac{\delta^k}{k!}\mathbb{E}H_k(w_t) = \sum_{k=2}^{\infty}\frac{h(\delta\delta_*^k)}{\delta_*\sqrt{n}k!} = n^{-1/2}h\left(\mathsf{e}^{\delta\delta_*} - 1 - \delta\delta_*\right)/\delta_*$$

which also goes to 0.

Applying Slutsky's Lemma to these factors and terms yields the needed result that

$$\sum_{k=3}^{\infty}\frac{\delta^{k-1}}{n(k-1)!}\sum_{t=1}^{n}\sqrt{n}\,(\widehat{v}_t - w_t)\,H_{k-1}(w_t) \xrightarrow{\mathbb{P}} 0.$$

and the second term in the sum is of smaller order. Thus proving the lemma.

References

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