

Implementation of Shi's non-degenerate Vuong test for non-nested models

Yves Croissant

1 The original Vuong test

Vuong (1989) proposed a test for non-nested model. He considers two competing models, $F_\beta = \{f(y|z; \beta); \beta \in \beta\}$ and $G_\gamma = \{g(y|z; \gamma); \gamma \in \Gamma\}$. Denoting $h(y|z)$ the true conditional density. The distance of F_β from the true model is measured by the minimum KLIC:

$$D_f = \mathbb{E}^0 [\ln h(y | z)] - \mathbb{E}^0 [\ln f(y | z; \beta_*)]$$

where \mathbb{E}^0 is the expected value using the true joint distribution of (y, X) and β_* is the pseudo-true value of β^1 . As the true model is unobserved, denoting $\theta^\top = (\beta^\top, \gamma^\top)$, we consider the difference of the KLIC distance to the true model of model G_γ and model F_β :

$$\Lambda(\theta) = D_g - D_f = \mathbb{E}^0 [\ln f(y | z; \beta_*)] - \mathbb{E}^0 [\ln g(y | z; \gamma_*)] = \mathbb{E}^0 \left[\ln \frac{f(y | z; \beta_*)}{g(y | z; \gamma_*)} \right]$$

The null hypothesis is that the distance of the two models to the true models are equal or, equivalently, that: $\Lambda = 0$. The alternative hypothesis is either $\Lambda > 0$, which means that F_β is better than G_γ or $\Lambda < 0$, which means that G_γ is better than F_β . Denoting, for a given random sample of size N , $\hat{\beta}$ and $\hat{\gamma}$ the maximum likelihood estimators of the two models and $\ln L_f(\hat{\beta})$ and $\ln L_g(\hat{\gamma})$ the maximum value of the log-likelihood functions of respectively F_β and G_γ , Λ can be consistently estimated by:

$$\hat{\Lambda}_N = \frac{1}{N} \sum_{n=1}^N \left(\ln f(y_n | x_n, \hat{\beta}) - \ln g(y_n | x_n, \hat{\gamma}) \right) = \frac{1}{N} \left(\ln L_f(\hat{\beta}) - \ln L_g(\hat{\gamma}) \right)$$

which is the likelihood ratio divided by the sample size. Note that the statistic of the standard likelihood ratio test, suitable for nested models is $2 \left(\ln L_f(\hat{\beta}) - \ln L_g(\hat{\gamma}) \right)$, which is $2N\hat{\Lambda}_N$.

The variance of Λ is:

$$\omega_*^2 = \mathbb{V}^o \left[\ln \frac{f(y | x; \beta_*)}{g(y | x; \gamma_*)} \right]$$

which can be consistently estimated by:

$$\hat{\omega}_N^2 = \frac{1}{N} \sum_{n=1}^N \left(\ln f(y_n | x_n, \hat{\beta}) - \ln g(y_n | x_n, \hat{\gamma}) \right)^2 - \hat{\Lambda}_N^2$$

Three different cases should be considered:

¹ β_* is called the pseudo-true value because f may be an uncorrect model.

- when the two models are nested, ω_*^2 is necessarily 0,
- when the two models are overlapping (which means than the models coincides for some values of the parameters), ω_*^2 may be equal to 0 or not,
- when the two models are strictly non-nested, ω_*^2 is necessarily strictly positive.

The distribution of the statistic depends on whether ω_*^2 is zero or positive. If ω_*^2 is positive, the statistic is $\hat{T}_N = \sqrt{N} \frac{\hat{\Lambda}_N}{\omega_N}$ and, under the null hypothesis that the two models are equivalent, follows a standard normal distribution. This is the case for two strictly non-nested models.

On the contrary, if $\omega_*^2 = 0$, the distribution is much more complicated. We need to define two matrices: A contains the expected values of the second derivatives of Λ :

$$A(\theta_*) = \mathbb{E}^0 \left[\frac{\partial^2 \Lambda}{\partial \theta \partial \theta^\top} \right] = \mathbb{E}^0 \left[\begin{array}{cc} \frac{\partial^2 \ln f}{\partial \beta \partial \beta^\top} & 0 \\ 0 & -\frac{\partial^2 \ln g}{\partial \gamma \partial \gamma^\top} \end{array} \right] = \left[\begin{array}{cc} A_f(\beta_*) & 0 \\ 0 & -A_g(\gamma_*) \end{array} \right]$$

and B the variance of its first derivatives:

$$B(\theta_*) = \mathbb{E}^0 \left[\frac{\partial \Lambda}{\partial \theta} \frac{\partial \Lambda}{\partial \theta^\top} \right] = \mathbb{E}^0 \left[\left(\frac{\partial \ln f}{\partial \beta}, -\frac{\partial \ln g}{\partial \gamma} \right) \left(\frac{\partial \ln f}{\partial \beta^\top}, -\frac{\partial \ln g}{\partial \gamma^\top} \right) \right] = \mathbb{E}^0 \left[\begin{array}{cc} \frac{\partial \ln f}{\partial \beta} \frac{\partial \ln f}{\partial \beta^\top} & -\frac{\partial \ln f}{\partial \beta} \frac{\partial \ln g}{\partial \gamma^\top} \\ -\frac{\partial \ln g}{\partial \gamma} \frac{\partial \ln f}{\partial \beta^\top} & \frac{\partial \ln g}{\partial \gamma} \frac{\partial \ln g}{\partial \gamma^\top} \end{array} \right]$$

or:

$$B(\theta_*) = \left[\begin{array}{cc} B_f(\beta_*) & -B_{fg}(\beta_*, \gamma_*) \\ -B_{gf}(\beta_*, \gamma_*) & B_g(\gamma_*) \end{array} \right]$$

Then:

$$W(\theta_*) = B(\theta_*) [-A(\theta_*)]^{-1} = \left[\begin{array}{cc} -B_f(\beta_*) A_f^{-1}(\beta_*) & -B_{fg}(\beta_*, \gamma_*) A_g^{-1}(\gamma_*) \\ B_{gf}(\gamma_*, \beta_*) A_f^{-1}(\beta_*) & B_g(\gamma_*) A_g^{-1}(\gamma_*) \end{array} \right]$$

Denote λ_* the eigen values of W . When $\omega_*^2 = 0$ (which is always the case for nested models), the statistic is the one used in the standard likelihood ratio test: $2(\ln L_f - \ln L_g) = 2N \hat{\Lambda}_N$ which, under the null, follows a weighted χ^2 distribution with weights equal to λ_* . The Vuong test can be seen in this context as a more robust version of the standard likelihood ratio test, because it doesn't assume, under the null, that the larger model is correctly specified.

Note that, if the larger model is correctly specified, the information matrix equality implies that $B_f(\theta_*) - A_f(\theta_*)$. In this case, the two matrices on the diagonal of W reduce to $-I_{K_f}$ and I_{K_g} , the trace of W to $K_g - K_f$ and the distribution of the statistic under the null reduce to a χ^2 with $K_g - K_f$ degrees of freedom.

The W matrix can be consistently estimate by computing the first and the second derivatives of the likelihood functions of the two models for $\hat{\theta}$. For example,

$$\hat{A}_f(\hat{\beta}) = \frac{1}{N} \sum_{n=1}^N \frac{\partial^2 \ln f}{\partial \beta \partial \beta^\top}(\hat{\beta}, x_n, y_n)$$

$$\hat{B}_{fg}(\hat{\theta}) = \frac{1}{N} \sum_{n=1}^N \frac{\partial \ln f}{\partial \beta}(\hat{\beta}, x_n, y_n) \frac{\partial \ln g}{\partial \gamma^\top}(\hat{\gamma}, x_n, y_n)$$

For the overlapping case, the test should be performed in two steps:

- the first step consists on testing whether ω_*^* is 0 or not. This hypothesis is based on the statistic $N\hat{\omega}^2$ which, under the null ($\omega_*^2 = 0$) follows a weighted χ^2 distributions with weights equal to λ_*^2 . If the null hypothesis is not rejected, the test stops at this step and the conclusion is that the two models are equivalent,
- if the null hypothesis is reject, the second step consists on applying the test for non-nested models previously described.

2 The non-degenerate Vuong test

Shi (2015) proposed a non-degenerate version of the Vuong (1989) test. She shows that the Vuong test has size distortion, leading to subsequent overrejection. The cause of this problem is that the distribution of $\hat{\Lambda}$ is discontinuous in the ω^2 parameter (namely a normal distribution if $\omega^2 > 0$ and a distribution related to a weight χ^2 distribution if $\omega^2 = 0$). Especially in small samples, it may be difficult to distinguish a positive versus a zero value of ω^2 because of sampling error. To solve this problem, using local asymptotic theory, Shi (2015) showed that, rewriting the Vuong statistic as:

$$\hat{T} = \frac{N\hat{\Lambda}_N}{\sqrt{N\hat{\omega}_N^2}}$$

the asymptotic distribution of the numerator and of the square of the denominator of the Vuong statistic is the same as:

$$\begin{pmatrix} N\hat{\Lambda}_N \\ N\hat{\omega}_N^2 \end{pmatrix} \rightarrow^d \begin{pmatrix} J_\Lambda \\ J_\omega \end{pmatrix} = \begin{pmatrix} \sigma z_\omega - z_\theta^\top V z_\theta / 2 \\ \sigma^2 - 2\sigma \rho_*^\top V z_\theta + z_\theta^\top V^2 z_\theta \end{pmatrix}$$

where:

$$\begin{pmatrix} z_\omega \\ z_\theta \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho_*^\top \\ \rho_* & I \end{pmatrix}\right),$$

ρ_* is a vector of length $K_f + K_g$, σ a positive scalar and V is the diagonal matrix containing the eigen values of $B^{\frac{1}{2}}A^{-1}B^{\frac{1}{2}}$.

Based on this result, Shi (2015) showed:

- that the expected value of the numerator is $-\text{trace}(V)/2$, the classical Vuong statistic is therefore biased and this bias can be severe in small samples and when the degree of parametrization of the two models are very different²,
- that the denominator, being random, can take values close to zero with a significant probability, which can generate fat tails in the distribution of the statistic.

Shi (2015) therefore proposed to modify the numerator of the Vuong statistic:

$$\hat{\Lambda}_N^{\text{mod}} = \hat{\Lambda}_N + \frac{\text{tr}(V)}{2N}$$

and to add a constant to the denominator, so that:

$$\left(\hat{\omega}^{\text{mod}}(c)\right)^2 = \hat{\omega}^2 + c \text{tr}(V)^2/N$$

²As the trace of V is the same as the trace of $A^{-1}B$, when the information matrix identity holds, it is equal to $-K_f + K_g$. The bias of the numerator is therefore caused by the difference in the degree of parametrization of the two models.

The non-degenerate Vuong test is then:

$$T_N^{\text{mod}} = \frac{\hat{\Lambda}_N^{\text{mod}}}{\hat{\omega}^{\text{mod}}} = \sqrt{N} \frac{\hat{\Lambda}_N + \text{tr}(V)/2N}{\sqrt{\hat{\omega}^2 + c \text{tr}(V)^2/N}}$$

The distribution of the modified Vuong statistic can be estimated by simulations: drawing in the distribution of $(z_\omega, z_\theta^\top)$, we compute for every draw J_Λ , J_ω and $J_\Lambda/\sqrt{J_\omega}$. As σ and ρ_* can't be estimated consistently, the supremum over these parameters are taken, and Shi (2015) indicates that ρ_* should be in this case a vector where all the elements are zero except for the one that coincides with the highest absolute value of V which is set to 1.

The Shi test is then computed as follow:

1. start with a given size for the test, say $\alpha = 0.05$,
2. for a given value of c , choose σ which maximize the simulated critical value for c and α ,
3. adjust c so that this critical value equals the normal critical value, up to a small discrepancy (say 0.1); for example, if the size is 5%, the target is $v_{1-\alpha/2} = 1.96 + 0.1 = 2.06$,
4. compute \hat{T}_N^{mod} for the given values of c and σ ; if $\hat{T}_N^{\text{mod}} > v_{1-\alpha/2}$, reject the null hypothesis at the α level,
5. to get a p-value, if $\hat{T}_N^{\text{mod}} > v_{1-\alpha/2}$ increase α and repeat the previous steps until a new value of α is obtained so that $\hat{T}_N^{\text{mod}} = v_{1-\alpha^* / 2}$, α^* being the p-value of the test.

3 Simulations

Shi (2015) provides an example of simulations of non-nested linear models that shows that the distribution of the Vuong statistic can be very different from a standard normal. The data generating process used for the simulations is:

$$y = 1 + \sum_{k=1}^{K_f} z_k^f + \sum_{k=1}^{K_g} z_k^g + \epsilon$$

where z^f is the set of K_f covariates that are used in the first model and z^g the set of K_g covariates used in the second model and $\epsilon \sim N(0, 1 - a^2)$. $z_k^f \sim N(0, a/\sqrt{K_f})$ and $z_k^g \sim N(0, a/\sqrt{K_g})$, so that the explained variance explained by the two competing models is the same (equal to a^2) and the null hypothesis of the Vuong test is true. The `sim_lm` unable to simulate values of the Vuong test. As in Shi (2015), we use a very different degree of parametrization for the two models, with $K_f = 15$ and $K_g = 1$.

```
Vuong <- ndvtest::sim_lm(N = 100, R = 1000, Kf = 15, Kg = 1, a = 0.5)
head(Vuong)
```

```
## [1] 3.6453881 1.9397603 1.2177819 0.6286946 2.0509829 2.5688364
```

```
mean(Vuong)
```

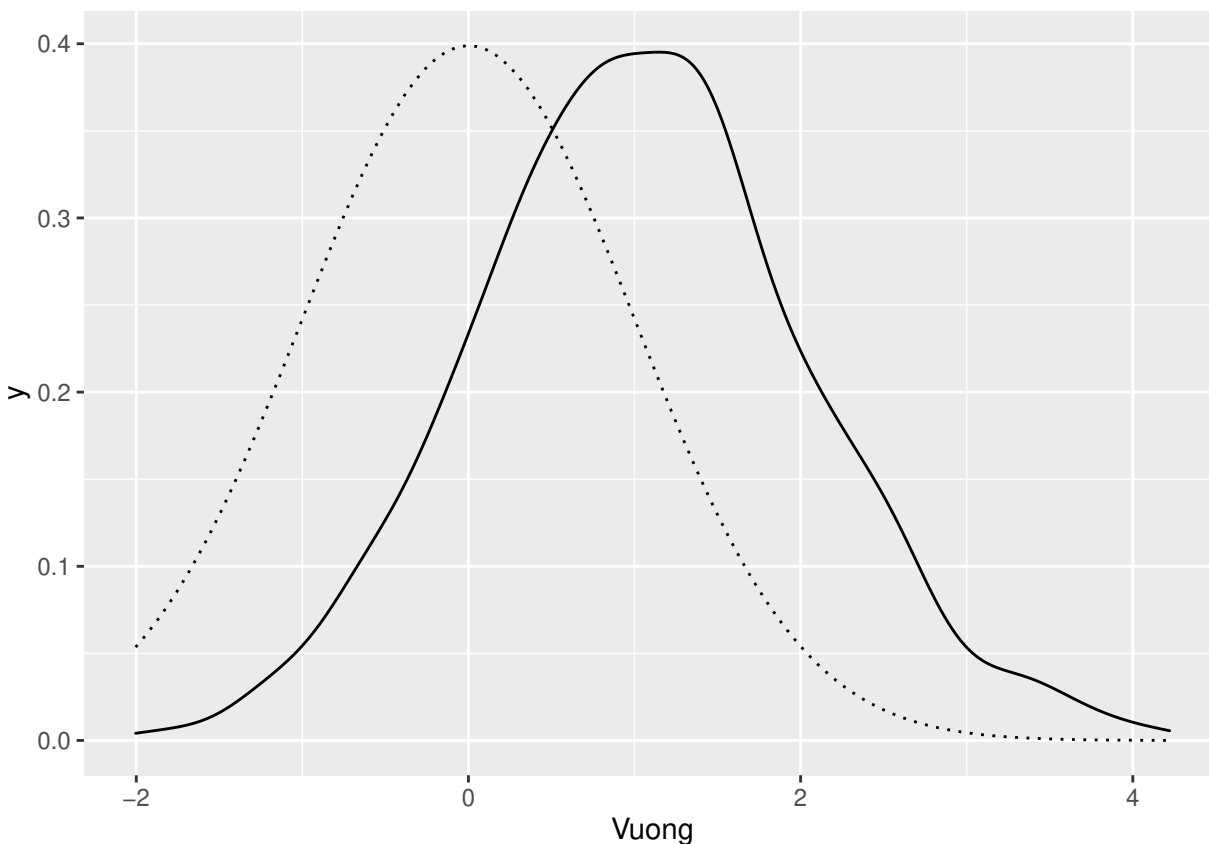
```
## [1] 1.036417
```

```
mean(abs(Vuong) > 1.96)
```

```
## [1] 0.176
```

We can see that the the mean of the statistic for the 1000 replications is far away from 0, which means that the numerator of the Vuong statistic is seriously biased. 17.6% of the values of the statistic are greater than the critical value so that the Vuong test will lead in such context a noticeable overrejection. The empirical pdf is shown in the following figure, along with the normal pdf.

```
library("ggplot2")
ggplot(data = data.frame(Vuong = Vuong)) + geom_density(aes(x = Vuong)) +
  geom_function(fun = dnorm, linetype = "dotted")
```



4 Implementation of the non-degenerate Vuong test

The `nonnest2` package provides a `vuongtest` function that implements the classical Vuong test. It has a `nest` argument (that is `FALSE` by default but can be set to `TRUE` to get the nested version of the Vuong test). This package also provide a `llcont` generic which returns a vector of length N containing the contribution of every observation to the log-likelihood.

The `ndvtest` package provides the `ndvtest` function. As for the `vuongtest` function, the two main arguments are two fitted models (say `model1` and `model2`). The $\hat{\Lambda}_n$ vector is obtained using `llcont(model1) - llcont(model2)`. The relevant matrices A_i and B_i are computed from the fitted models using the `estfun` and the `meat` functions from the `sandwich` package. More precisely, A^{-1} is `bdiag(-bread(model1), bread(model2))3` and B is `crossprod(estfun(model1), - estfun(model2)) / N`, where N is the sample size. Therefore, the `ndvtest` function can be used with any models for which a `llcont`, a `estfun` and a `bread` method is available.

³`bdiag` is a function that construct a block-diagonal matrix from its arguments.

5 Applications

5.1 Voter turnout

The first application is the example used in Shi (2015) and is used to compare our **R** program with Shi's **stata**'s program. Coate and Conlin (2004) used several models of electoral participation, using data concerning referenda about alcohol sales regulation in Texas. Three models are estimated: the preferred group-utilitarian model, a "simple, but plausible, alternative: the intensity model" and a reduced form model estimated by the seemingly unrelated residuals method. They are provided in the **ndvtest** package as **turnout**, a list of three fitted models⁴. The results of the Shi test are given below. We first compute the Shi statistic for an error level of 5%. We therefore set the **size** argument to 0.05 (this is actually the default value) and the **pval** argument to **FALSE**.

```
library("ndvtest")
test <- ndvtest(turnout$group, turnout$intens, size = 0.05, pval = FALSE)
test

##
## Non-degenerate Vuong test for non-nested models
##
## data: turnout$group-turnout$intens
## z = 1.7441, size = 0.050000, vuong_stat = 2.084529, constant =
## 0.636510, crit-value = 2.059733, sum e.v. = -11.043994, vuong_p.value =
## 0.037112
## alternative hypothesis: different models
```

The Shi statistic is 1.744, which is smaller than the critical value 2.06. Therefore, based on the Shi test, we can't reject the hypothesis that the two competing models are equivalent at the 5% level. The value of the constant c is also reported, as is the sum of the eigen values of the V matrix (**sum e.v.**). The classical Vuong statistic is also reported (2.085) and is greater than the 5% normal critical value (the p-value is 0.037). Therefore, the classical Vuong test and the non-degenerate version lead to opposite conclusions at the 5% level.

To get only the classical Vuong test, the **nd** argument can be set to **FALSE**:

```
ndvtest(turnout$group, turnout$intens, nd = FALSE)

##
## Vuong test for non-nested models
##
## data: turnout$group-turnout$intens
## z = 2.0845, p-value = 0.03711
## alternative hypothesis: different models
```

To get the p-value of the non-degenerate Vuong test, the **pval** argument should be set to **TRUE**.

```
test <- ndvtest(turnout$group, turnout$intens, pval = TRUE)
test
```

⁴The estimation is rather complicated because some linear constraints are used to compute the maximum likelihood estimator in Coate and Conlin (2004)'s **stata** script. This is the reason why we provide only the results of the estimations, performed using the **maxLik** package.

```
##
## Non-degenerate Vuong test for non-nested models
##
## data: turnout$group-turnout$intens
## z = 1.7976, vuong_stat = 2.084529, constant = 0.124584, sum e.v. =
## -11.043994, vuong_p.value = 0.037112, p-value = 0.0896
## alternative hypothesis: different models
```

The results indicate that the p-value is 0.09, which confirms that the Shi test concludes that the two model are equivalent at the 5% level.

5.2 The demand for medical care (stricly non-nested models)

The second example compares different count models, using data on demand for medical care. These data were used by Deb and Trivedi (1997) and by Zeileis, Kleiber, and Jackman (2008). They are available as `AER::NMES1988`. The response is the number of physician office visits (`visits`) and the covariates are the number of hospital stays (`hosp`), a factor indicating self-perceived health status (`health`), the number of chronic conditions (`chronic`), the gender (`gender`), the number of years of education (`school`) and a factor indicating whether the individual is covered by private insurance (`insurance`).

The poisson model is often two simple as the poisson distribution hypothesis may be rejected for two reasons:

- the first one is that it imposes the equality of the conditional expectation and variance and that, with real data, we often observe that the variance is much larger than the expectation,
- the second one is that the probability of 0 for real data are often much larger that the one implied by the poisson distribution.

We'll deal here only with the second problem. Two competing models can be used:

- the zero-inflated (or zip) model, as it names suggests, inflates the probability of zero, compared to the Poisson model,
- the hurdle model is a two-part model, with a binomial model which explains the probability of 0 and a zero-truncated poisson model that explains the probabilities of positive values.

These two models can be estimated using respectively `pscl::zeroinfl` and `pscl::hurdle`.

```
data("NMES1988", package = "AER")
zi <- pscl::zeroinfl(visits ~ hospital + health + chronic + gender + school + insurance,
                    data = NMES1988)
hl <- pscl::hurdle(visits ~ hospital + health + chronic + gender + school + insurance,
                  data = NMES1988)
```

We first use two previous implementation of the classical Vuong test. The first one is `pscl::vuong`.

```
pscl::vuong(zi, hl)

## Vuong Non-Nested Hypothesis Test-Statistic:
## (test-statistic is asymptotically distributed N(0,1) under the
## null that the models are indistinguishable)
## -----
##              Vuong z-statistic              H_A  p-value
## Raw                2.189664 model1 > model2 0.014274
## AIC-corrected      2.189664 model1 > model2 0.014274
## BIC-corrected      2.189664 model1 > model2 0.014274
```

The alternative hypothesis is written as the first model being better than the second one, the p-value is therefore one-sided. The null hypothesis is therefore rejected at the 5% level (p-value = 0.0143), and it's also the case if the alternative were written as one model being better than the other one (in this case, the p-value would be $0.0143 \times 2 = 0.0286$).

An other implementation of the classical Vuong test is `nonnest2::vuong`.

```
nonnest2::vuongtest(zi, hl)

##
## Model 1
## Class: zeroinfl
## Call: pscl::zeroinfl(formula = visits ~ hospital + health + chronic + ...
##
## Model 2
## Class: hurdle
## Call: pscl::hurdle(formula = visits ~ hospital + health + chronic + ...
##
## Variance test
## H0: Model 1 and Model 2 are indistinguishable
## H1: Model 1 and Model 2 are distinguishable
## w2 = 0.000, p = 0.0124
##
## Non-nested likelihood ratio test
## H0: Model fits are equal for the focal population
## H1A: Model 1 fits better than Model 2
## z = 2.190, p = 0.0143
## H1B: Model 2 fits better than Model 1
## z = 2.190, p = 0.9857
```

This implementation of the test proceeds in two steps. The first step is the variance test, which is normally only relevant when the two models are overlapping. The hypothesis that $\omega^2 = 0$ is rejected at the 5% level (p-value of 0.0124), so that the second step is performed. Two alternatives hypothesis are used, the first one being that the first model is better than the second model and the second one being the opposite. Therefore, one sided p-values are used and we get the same conclusion than previously that the zero-inflated Poisson model is better than the hurdle model at the 5% level.

We then compute the non-degenerate Vuong test:

```
ndvttest(zi, hl)

##
## Non-degenerate Vuong test for non-nested models
##
## data: zi-hl
## z = 2.0005, vuong_stat = 2.189912, constant = 0.486841, sum e.v. =
## -0.010899, vuong_p.value = 0.028531, p-value = 0.0573
## alternative hypothesis: different models
```

The null hypothesis that the two models are equivalent is not rejected using the non-degenerate Vuong test, as the p-value is now equal to 0.057.

5.3 Transport mode choice (nested models)

The third example concerns transport mode choice in Canada. The dataset, provided by the `mlogit` package is called `ModeCanada` and has been used extensively in the transport demand literature (see in particular Bhat 1995; Koppelman and Wen 2000; and Wen and Koppelman 2001). The following example is from Croissant (2020). The raw data set is first transformed to make it suitable for the estimation of discrete choice models. The sample is restricted to the individuals for which 4 transport modes are available (bus, air, train and car).

```
library("mlogit")

## Le chargement a nécessité le package : dfixd
##
## Attachement du package : 'dfixd'
##
## L'objet suivant est masqué depuis 'package:stats':
##
##      filter

data("ModeCanada")
MC <- mlogit.data(ModeCanada, subset = noalt == 4, chid.var = "case",
                 alt.var = "alt", drop.index = TRUE)
```

We first estimate the simplest discrete choice model, which is the multinomial logit model. The bus share being negligible, the choice set is restricted to the three other modes and the reference mode is set to car.

```
ml <- mlogit(choice ~ freq + cost + ivt + ovt | urban + income, MC,
            reflevel = 'car', alt.subset = c("car", "train", "air"))
```

This model relies on the hypothesis that the unobserved component of the utility functions for the different modes are independent and identical Gumbell variables. Bhat (1995) proposed the heteroscedastic logit for which the errors follow a general Gumbell distributions with a supplementary scale parameter to be estimated. As the overall scale of utility is not identified, the scale parameter of the reference alternative (car) is set to one.

```
hl <- mlogit(choice ~ freq + cost + ivt + ovt | urban + income, MC,
            reflevel = 'car', alt.subset = c("car", "train", "air"),
            heterosc = TRUE)
coef(summary(hl))
```

```
##              Estimate Std. Error  z-value  Pr(>|z|)
## (Intercept):train  0.678393435 0.332762598  2.038671 4.148288e-02
## (Intercept):air   0.656754399 0.468163091  1.402832 1.606668e-01
## freq              0.063924677 0.004916769 13.001360 0.000000e+00
## cost              -0.026961457 0.004283139 -6.294789 3.078178e-10
## ivt               -0.009680773 0.001053874 -9.185892 0.000000e+00
## ovt               -0.032165526 0.003593007 -8.952258 0.000000e+00
## urban:train       0.797131578 0.120739176  6.602096 4.053868e-11
## urban:air         0.445472634 0.082160945  5.421951 5.895197e-08
## income:train     -0.012597857 0.003994180 -3.154053 1.610196e-03
## income:air        0.018859983 0.003215926  5.864558 4.503294e-09
## sp.train          1.237182865 0.110460959 11.200182 0.000000e+00
## sp.air            0.540323852 0.111835294  4.831425 1.355592e-06
```

The two supplementary coefficients are `sp.train` and `sp.air`. The student statistics reported are irrelevant because they test the hypothesis that these parameters are 0, as the relevant hypothesis of homoscedasticity is that both of them equal one. The heteroscedastic logit being nested in the multinomial logit model, we can first use the three classical tests: the Wald test (based on the unconstrained model `hl`), the score test (based on the constrained model `m1`) and the likelihood ratio model (based on the comparison of both models).

To perform the Wald test, we use `lmtest::waldtest`, for which a special method is provided by the `mlogit` package. The arguments are the unconstrained model (`hl`) and the update that should be used in order to get the constrained model (`heterosc = FALSE`). To compute the scoretest, we use `mlogit::scoretest`, for which the arguments are the constrained model (`m1`) and the update that should be used in order to get the unconstrained model (`heterosc = TRUE`). Finally, the likelihood ratio test is performed using `lmtest::lrtest`.

```
lmtest::waldtest(hl, heterosc = FALSE)
```

```
##
## Wald test
##
## data: homoscedasticity
## chisq = 25.196, df = 2, p-value = 3.38e-06
```

```
scoretest(m1, heterosc = TRUE)
```

```
##
## score test
##
## data: heterosc = TRUE
## chisq = 9.4883, df = 2, p-value = 0.008703
## alternative hypothesis: heteroscedastic model
```

```
lmtest::lrtest(hl, m1)
```

```
## Likelihood ratio test
##
## Model 1: choice ~ freq + cost + ivt + ovt | urban + income
## Model 2: choice ~ freq + cost + ivt + ovt | urban + income
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 12 -1838.1
## 2 10 -1841.6 -2 6.8882 0.03193 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The three statistics are χ^2 with two degrees of freedom under the null hypothesis of homoscedasticity. The three tests reject the null hypothesis at the 5% level, and even at the 1% level for the Wald and the score test. These three tests relies on the hypothesis that, under the null, the constrained model is the true model. We can get rid of this hypothesis using a Vuong test. Note the use of the `nested` argument that is set to `TRUE`:

```
ndvtest(hl, m1, nested = TRUE)
```

```
##
## Non-degenerate Vuong test for nested models
```

```
##
## data:  hl-ml
## z = 0.4554, vuong_stat = 6.888241, vuong_p.value = 0.047927, p-value =
## 0.211
## alternative hypothesis: different models
```

The homoscedasticity hypothesis is still rejected at the 5% level for the classical Vuong test (the p-value is 0.048), but it is not using the non-degenerate Vuong test (p-value of 0.211).

5.4 Transport mode choice (overlapping models)

We consider finally another dataset from `mlogit` called `RiskyTransport`, that has been used by Leon and Miguel (2017) and concerns the choice of one mode (among water-taxi, ferry, hovercraft and helicopter) for trips from Sierra Leone's international airport to downtown Freetown.

```
library("mlogit")
data("RiskyTransport")
RT <- mlogit.data(RiskyTransport, shape = "long", choice = "choice",
                 chid.var = "chid", alt.var = "mode", id.var = "id")
```

We estimate models with only one covariate, the generalized cost of the mode. We estimate three models: the basic multinomial logit model, the heteroscedastic model and a nested model where alternatives are grouped in two nests according to the fact that they are fast or slow modes.

```
m1 <- mlogit(choice ~ cost, data = RT)
hl <- mlogit(choice ~ cost, data = RT, heterosc = TRUE)
nl <- mlogit(formula = choice ~ cost, data = RT,
             nests = list(fast = c("Helicopter", "Hovercraft"),
                          slow = c("WaterTaxi", "Ferry")),
             un.nest.el = TRUE)
modelsummary::msummary(list(multinomial = m1, heteroscedastic = hl, nested = nl))
```

Compared to the multinomial model, the heteroscedastic model has 3 supplementary coefficients (the scale parameters for 3 modes, the one for the reference mode being set to 1) and the nested logit model has one supplementary parameter which is the nest elasticity (`iv` in the table). Both models reduce to the multinomial logit model if:

- `sp.WaterTaxi = sp.Ferry = sp.Hovercraft = 1` for the heteroscedastic model,
- `iv = 1` for the nested logit model.

Therefore, the two models are over-lapping, as they reduce to the same model (the multinomial logit model) for some values of the parameters.

The first step of the test is the variance test. It can be performed using `ndvtest` by setting the argument `vartest` to `TRUE`:

```
ndvtest(nl, hl, vartest = TRUE)

##
## variance test
##
## data:  nl-hl
## w2 = 0.047327, p-value < 2.2e-16
## alternative hypothesis: positive variance
```

	multinomial	heteroscedastic	nested
(Intercept) × WaterTaxi	1.754 (0.159)	1.955 (0.224)	1.581 (0.127)
(Intercept) × Ferry	1.445 (0.168)	0.417 (0.401)	1.359 (0.118)
(Intercept) × Hovercraft	0.844 (0.167)	-3.666 (2.511)	0.461 (0.148)
cost	-0.009 (0.001)	-0.022 (0.004)	-0.007 (0.001)
sp.WaterTaxi		0.745 (0.523)	
sp.Ferry		2.840 (0.547)	
sp.Hovercraft		4.551 (1.761)	
iv			0.554 (0.092)
Num.Obs.	1793	1793	1793
AIC	3344.2	3300.7	3330.4
BIC			
Log.Lik.	-1668.108	-1643.348	-1660.198
rho2	0.151	0.163	0.155
rho20	0.329	0.339	0.332

The null hypothesis that $\omega^2 = 0$ is rejected. We can then proceed to the second step, which is the test for non-nested models.

```
ndvtest(h1, n1)
```

```
##
## Non-degenerate Vuong test for non-nested models
##
## data: h1-n1
## z = 1.7298, vuong_stat = 1.829208, constant = 0.000000, sum e.v. =
## -1.832021, vuong_p.value = 0.067369, p-value = 0.0975
## alternative hypothesis: different models
```

The classical and the non-degenerate Vuong tests both conclude that the two models are equivalent at the 5% level, but that the heteroscedastic model is better than the nested logit model at the 10% level.

References

- Bhat, Chandra R. 1995. “A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice.” *Transportation Research Part B: Methodological* 29 (6): 471–83. <https://www.sciencedirect.com/science/article/pii/0191261595000156>.
- Coate, Stephen, and Michael Conlin. 2004. “A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence.” *American Economic Review* 94 (5): 1476–1504. <https://doi.org/10.1257/0002828043052231>.
- Croissant, Yves. 2020. “Estimation of Random Utility Models in R: The Mlogit Package.” *Journal of Statistical Software* 95 (11): 1–41. <https://doi.org/10.18637/jss.v095.i11>.

- Deb, Partha, and Pravin K. Trivedi. 1997. "Demand for Medical Care by the Elderly: A Finite Mixture Approach." *Journal of Applied Econometrics* 12 (3): 313–36. [https://doi.org/https://doi.org/10.1002/\(SICI\)1099-1255\(199705\)12:3<313::AID-JAE440>3.0.CO;2-G](https://doi.org/https://doi.org/10.1002/(SICI)1099-1255(199705)12:3<313::AID-JAE440>3.0.CO;2-G).
- Koppelman, Frank S., and Chieh-Hua Wen. 2000. "The Paired Combinatorial Logit Model: Properties, Estimation and Application." *Transportation Research Part B: Methodological* 34 (2): 75–89. <https://www.sciencedirect.com/science/article/pii/S0191261599000120>.
- Leon, Gianmarco, and Edward Miguel. 2017. "Risky Transportation Choices and the Value of a Statistical Life." *American Economic Journal: Applied Economics* 9 (1): 202–28.
- Shi, Xiaoxia. 2015. "A Nondegenerate Vuong Test." *Quantitative Economics*, 85–121.
- Vuong, Quang H. 1989. "Likelihood Ratio Tests for Selection and Non-Nested Hypotheses." *Econometrica* 57 (2): 397–33.
- Wen, Chieh-Hua, and Frank S Koppelman. 2001. "The Generalized Nested Logit Model." *Transportation Research Part B: Methodological* 35 (7): 627–41. <https://www.sciencedirect.com/science/article/pii/S019126150000045X>.
- Zeileis, Achim, Christian Kleiber, and Simon Jackman. 2008. "Regression Models for Count Data in R." *Journal of Statistical Software* 27 (8): 1–25. <https://doi.org/10.18637/jss.v027.i08>.