

Kenneth Train's exercises using the `mlogit` package for R

Kenneth Train

University of California-Berkeley

Yves Croissant

Université de la Réunion

Abstract

This text presents the answers to the exercises provided by Kenneth Train on his web site using the `mlogit` package for R.

Keywords: discrete choice models, maximum likelihood estimation, R, econometrics.

1. Multinomial logit model

1. The problem set uses data on choice of heating system in California houses. The data set `Heating` from the `mlogit` package contains the data in R format. The observations consist of single-family houses in California that were newly built and had central air-conditioning. The choice is among heating systems. Five types of systems are considered to have been possible:

1. gas central (`gc`),
2. gas room (`gr`),
3. electric central (`ec`),
4. electric room (`er`),
5. heat pump (`hp`).

There are 900 observations with the following variables:

- `idcase` gives the observation number (1-900),
- `depvar` identifies the chosen alternative (`gc`, `gr`, `ec`, `er`, `hp`),
- `ic.alt` is the installation cost for the 5 alternatives,
- `oc.alt` is the annual operating cost for the 5 alternatives,
- `income` is the annual income of the household,

- `agehed` is the age of the household head,
- `rooms` is the number of rooms in the house,
- `region` a factor with levels `ncostl` (northern coastal region), `scostl` (southern coastal region), `mountn` (mountain region), `valley` (central valley region).

Note that the attributes of the alternatives, namely, installation cost and operating cost, take a different value for each alternative. Therefore, there are 5 installation costs (one for each of the 5 systems) and 5 operating costs. To estimate the logit model, the researcher needs data on the attributes of all the alternatives, not just the attributes for the chosen alternative. For example, it is not sufficient for the researcher to determine how much was paid for the system that was actually installed (i.e., the bill for the installation). The researcher needs to determine how much it would have cost to install each of the systems if they had been installed. The importance of costs in the choice process (i.e., the coefficients of installation and operating costs) is determined through comparison of the costs of the chosen system with the costs of the non-chosen systems.

For these data, the costs were calculated as the amount the system would cost if it were installed in the house, given the characteristics of the house (such as size), the price of gas and electricity in the house location, and the weather conditions in the area (which determine the necessary capacity of the system and the amount it will be run.) These cost are conditional on the house having central air-conditioning. (That's why the installation cost of gas central is lower than that for gas room: the central system can use the air-conditioning ducts that have been installed.)

In a logit model, each variable takes a different value in each alternative. So, in our case, for example, we want to know the coefficient of installation cost in the logit model of system choice. The variable installation cost in the model actually consists of five variables in the dataset: `ic.gc`, `ic.gr`, `ic.ec`, `ic.er` and `ic.hp`, for the installation costs of the five systems. In the current code, there are two variables in the logit model. The first variable is called `ic` for installation cost. This variable consists of five variables in the dataset: `ic.gc` in the first alternative, `ic.gr` in the second alternative, etc.

2. Run a model with installation cost and operating cost, without intercepts

(a) Do the estimated coefficients have the expected signs?

```
R> library("mlogit")
R> #source("~/Forge/mlogit/chargement.R")
R> data("Heating", package = "mlogit")
R> H <- mlogit.data(Heating, shape="wide", choice="depvar", varying=c(3:12))
R> m <- mlogit(depvar~ic+oc|0, H)
R> summary(m)
```

Call:

```
mlogit(formula = depvar ~ ic + oc | 0, data = H, method = "nr",
        print.level = 0)
```

Frequencies of alternatives:

```

      ec      er      gc      gr      hp
0.071111 0.093333 0.636667 0.143333 0.055556

```

nr method

4 iterations, 0h:0m:0s

g'(-H)^-1g = 1.56E-07

gradient close to zero

Coefficients :

```

      Estimate Std. Error t-value Pr(>|t|)
ic -0.00623187  0.00035277 -17.665 < 2.2e-16 ***
oc -0.00458008  0.00032216 -14.217 < 2.2e-16 ***
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -1095.2

Yes, they are negative as expected, meaning that as the cost of a system rises (and the costs of the other systems remain the same) the probability of that system being chosen falls.

(b) Are both coefficients significantly different from zero?

Yes, the t-statistics are greater than 1.96, which is the critical level for 95% confidence level.

(c) How closely do the average probabilities match the shares of customers choosing each alternative?

```
R> apply(fitted(m, outcome=FALSE), 2, mean)
```

```

      ec      er      gc      gr      hp
0.10413057 0.05141477 0.51695653 0.24030898 0.08718915

```

Not very well. 63.67% of the sample chose `gc` (as shown at the top of the summary) and yet the estimated model gives an average probability of only 51.695%. The other alternatives are also fairly poorly predicted. We will find how to fix this problem in one of the models below.

(d) The ratio of coefficients usually provides economically meaningful information. The willingness to pay (*wtp*) through higher installation cost for a one-dollar reduction in operating costs is the ratio of the operating cost coefficient to the installation cost coefficient. What is the estimated *wtp* from this model? Is it reasonable in magnitude?

$$U = \beta_{ic}ic + \beta_{oc}oc$$

$$dU = \beta_{ic}dic + \beta_{oc}doc = 0 \Rightarrow -\frac{dic}{doc} \Big|_{dU=0} = \frac{\beta_{oc}}{\beta_{ic}}$$

```
R> coef(m) ["oc"] / coef(m) ["ic"]
```

```

      oc
0.7349453

```

The model implies that the decision-maker is willing to pay \$.73 (ie., 73 cents) in higher installation cost in order to reduce annual operating costs by \$1.

A \$1 reduction in annual operating costs recurs each year. It is unreasonable to think that the decision-maker is only willing to pay only 73 cents as a one-time payment in return for a \$1/year stream of saving. This unreasonable implication is another reason (along with the inaccurate average probabilities) to believe this model is not so good. We will find below how the model can be improved.

(e) We can use the estimated wtp to obtain an estimate of the discount rate that is implied by the model of choice of operating system. The present value of the future operating costs is the discounted sum of operating costs over the life of the system: $PV = \sum_{t=1}^L \frac{OC}{(1+r)^t}$ where r is the discount rate and L being the life of the system. As L rises, the PV approaches OC/r . Therefore, for a system with a sufficiently long life (which we will assume these systems have), a one-dollar reduction in OC reduces the present value of future operating costs by $(1/r)$. This means that if the person choosing the system were incurring the installation costs and the operating costs over the life of the system, and rationally traded-off the two at a discount rate of r , the decisionmaker's wtp for operating cost reductions would be $(1/r)$. Given this, what value of r is implied by the estimated wtp that you calculated in part (c)? Is this reasonable?

$U = aLC$ where LC is lifecycle cost, equal to the sum of installation cost and the present value of operating costs: $LC = IC + (1/r)OC$. Substituting, we have $U = aIC + (a/r)OC$.

The models estimates a as -0.00623 and a/r as -0.00457 . So $r = a/(a/r) = -.000623/.00457 = 1.36$ or 136% discount rate. This is not reasonable, because it is far too high.

3. Estimate a model that imposes the constraint that $r = 0.12$ (such that $wtp = 8.33$). Test the hypothesis that $r = 0.12$.

To impose this constraint, we create a lifecycle cost that embodies the constraint $lcc = ic + oc/0.12$ and estimate the model with this variable.

```
R> H$lcc=H$ic+H$oc/0.12
R> mlcc <- mlogit(depvar~lcc|0, H)
R> lrtest(m, mlcc)
```

Likelihood ratio test

```
Model 1: depvar ~ ic + oc | 0
Model 2: depvar ~ lcc | 0
#Df LogLik Df Chisq Pr(>Chisq)
1 2 -1095.2
2 1 -1248.7 -1 306.93 < 2.2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R> qchisq(0.05, df = 1, lower.tail = FALSE)
```

```
[1] 3.841459
```

We perform a likelihood ratio test. The $\ln L$ for this constrained model is -1248.7 . The $\ln L$ for the unconstrained model is -1095.2 . The test statistic is twice the difference in $\ln L$: $2(1248.7 - 1095.2) = 307$. This test is for one restriction (ie a restriction on the relation of the coefficient of operating cost to that of installation cost.) We therefore compare 307 with the critical value of chi-squared with 1 degree of freedom. This critical value for 95% confidence is 3.8. Since the statistic exceeds the critical value, we reject the hypothesis that $r = 0.12$.

4. Add alternative-specific constants to the model. With J alternatives, at most $J - 1$ alternative-specific constants can be estimated. The coefficients of $J - 1$ constants are interpreted as relative to alternative J th alternative. Normalize the constant for the alternative `hp` to 0.

(a) How well do the estimated probabilities match the shares of customers choosing each alternative?

```
R> mc <- mlogit(depvar~ic+oc, H, reflevel = 'hp')
R> summary(mc)
```

Call:

```
mlogit(formula = depvar ~ ic + oc, data = H, reflevel = "hp",
        method = "nr", print.level = 0)
```

Frequencies of alternatives:

```
      hp      ec      er      gc      gr
0.055556 0.071111 0.093333 0.636667 0.143333
```

nr method

6 iterations, 0h:0m:0s

$g'(-H)^{-1}g = 9.58E-06$

successive function values within tolerance limits

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
ec:(intercept)	1.65884594	0.44841936	3.6993	0.0002162 ***
er:(intercept)	1.85343697	0.36195509	5.1206	3.045e-07 ***
gc:(intercept)	1.71097930	0.22674214	7.5459	4.485e-14 ***
gr:(intercept)	0.30826328	0.20659222	1.4921	0.1356640
ic	-0.00153315	0.00062086	-2.4694	0.0135333 *
oc	-0.00699637	0.00155408	-4.5019	6.734e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -1008.2

McFadden R^2 : 0.013691

Likelihood ratio test : $\text{chisq} = 27.99$ (p.value = $8.3572e-07$)

```
R> apply(fitted(mc, outcome = FALSE), 2, mean)
```

```
      hp      ec      er      gc      gr
0.05555556 0.07111111 0.09333333 0.63666667 0.14333333
```

Note that they match exactly: alternative-specific constants in a logit model insure that the average probabilities equal the observed shares.

(b) Calculate the *wtp* and discount rate *r* that is implied by the estimates. Are these reasonable?

```
R> wtp <- coef(mc)["oc"]/coef(mc)["ic"]
R> wtp
```

```
      oc
4.563385
```

```
R> r <- 1/wtp
R> r
```

```
      oc
0.2191356
```

The decision-maker is willing to pay \$4.56 for a \$1 year stream of savings. This implies $r = 0.22$. The decision-maker applies a 22% discount rate. These results are certainly more reasonable than in the previous model. The decision-maker is still estimated to be valuing saving somewhat less than would seem rational (ie applying a higher discount rate than seems reasonable). However, we need to remember that the decision-maker here is the builder. If home buyers were perfectly informed, then the builder would adopt the buyer's discount rate. However, the builder would adopt a higher discount rate if home buyers were not perfectly informed about (or believed) the stream of saving.

(c) This model contains constants for all alternatives *ec-er-gc-gr*, with the constant for alternative *hp* normalized to zero. Suppose you had included constants for alternatives *ec-er-gc-hp*, with the constant for alternative *gr* normalized to zero. What would be the estimated coefficient of the constant for alternative *gc*? Figure this out logically rather than actually estimating the model.

We know that when the *hp* is left out, the constant for alternative *gc* is 1.71074 meaning that the average impact of uncluded factors is 1.71074 higher for alternative *gc* than for alternative *hp*. Similarly, the constant for alternative *gr* is 0.30777. If *gr* were left out instead of *hp*, then all the constants would be relative to alternative *gr*. The constant for alternative *gc* would be $1.71074 - .30777 = 1.40297$. That is, the average impact of uncluded factors is 1.40297 higher for alt *gc* than alt *gr*. Similarly for the other alternatives. Note the the constant for alt 5 would be $0 - .30777 = -.30777$, since *hp* is normalized to zero in the model with *hp* left out.

```
R> update(mc, refllevel="gr")
```

Call:

```
mlogit(formula = depvar ~ ic + oc, data = H, refllevel = "gr", method = "nr", print.level =
```

Coefficients:

```
ec:(intercept)  er:(intercept)  gc:(intercept)  hp:(intercept)          ic
      1.3505827      1.5451737      1.4027160      -0.3082633      -0.0015332
              oc
      -0.0069964
```

5. Now try some models with sociodemographic variables entering.

(a) Enter installation cost divided by income, instead of installation cost. With this specification, the magnitude of the installation cost coefficient is inversely related to income, such that high income households are less concerned with installation costs than lower income households. Does dividing installation cost by income seem to make the model better or worse?

```
R> mi <- mlogit(depvar~oc+I(ic/income), H)
R> summary(mi)
```

Call:

```
mlogit(formula = depvar ~ oc + I(ic/income), data = H, method = "nr",
       print.level = 0)
```

Frequencies of alternatives:

```
      ec      er      gc      gr      hp
0.071111 0.093333 0.636667 0.143333 0.055556
```

nr method

6 iterations, 0h:0m:0s

$g'(-H)^{-1}g = 1.03E-05$

successive function values within tolerance limits

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
er:(intercept)	0.0639934	0.1944893	0.3290	0.742131
gc:(intercept)	0.0563481	0.4650251	0.1212	0.903555
gr:(intercept)	-1.4653063	0.5033845	-2.9109	0.003604 **
hp:(intercept)	-1.8700773	0.4364248	-4.2850	1.827e-05 ***
oc	-0.0071066	0.0015518	-4.5797	4.657e-06 ***
I(ic/income)	-0.0027658	0.0018944	-1.4600	0.144298

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -1010.2

McFadden R²: 0.011765

Likelihood ratio test : chisq = 24.052 (p.value = 5.9854e-06)

The model seems to get worse. The $\ln L$ is lower (more negative) and the coefficient on installation cost becomes insignificant (t-stat below 2).

(b) Instead of dividing installation cost by income, enter alternative-specific income effects. What do the estimates imply about the impact of income on the choice of central systems versus room system? Do these income terms enter significantly?

```
R> mi2 <- mlogit(depvar~oc+ic/income, H, reflevel="hp")
```

The model implies that as income rises, the probability of heat pump rises relative to all the others (since income in the heat pump alt is normalized to zero, and the others enter with negative signs

such that they are lower than that for heat pumps. Also, as income rises, the probability of gas room drops relative to the other non-heat-pump systems (since it is most negative).

Do these income terms enter significantly? No. It seems that income doesn't really have an effect. Maybe this is because income is for the family that lives in the house, whereas the builder made decision of which system to install.

```
R> lrtest(mc, mi2)
```

Likelihood ratio test

```
Model 1: depvar ~ ic + oc
Model 2: depvar ~ oc + ic | income
#Df  LogLik Df  Chisq Pr(>Chisq)
1    6 -1008.2
2   10 -1005.9  4  4.6803    0.3217
```

```
R> waldtest(mc, mi2)
```

Wald test

```
Model 1: depvar ~ ic + oc
Model 2: depvar ~ oc + ic | income
Res.Df Df  Chisq Pr(>Chisq)
1      894
2      890  4  4.6456    0.3256
```

```
R> scoretest(mc, mi2)
```

score test

```
data:  depvar ~ oc + ic | income
chisq = 4.6761, df = 4, p-value = 0.3222
alternative hypothesis: unconstrained model
```

(c) Try other models. Determine which model you think is best from these data.

I'm not going to give what I consider my best model: your ideas on what's best are what matter here.

6. We now are going to consider the use of logit model for prediction. Estimate a model with installation costs, operating costs, and alternative specific constants. Calculate the probabilities for each house explicitly. Check to be sure that the mean probabilities are the same as you got in exercise 4.

```
R> X <- model.matrix(mc)
R> alt <- index(H)$alt
R> chid <- index(H)$chid
R> eXb <- as.numeric(exp(X %*% coef(mc)))
R> SeXb <- tapply(eXb, chid, sum)
R> P <- eXb / SeXb[chid]
```

```
R> P <- matrix(P, ncol = 5, byrow = TRUE)
R> head(P)

      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.05107444 0.07035738 0.6329116 0.1877416 0.05791494
[2,] 0.04849337 0.06420595 0.6644519 0.1558322 0.06701658
[3,] 0.07440281 0.08716904 0.6387765 0.1439919 0.05565974
[4,] 0.07264503 0.11879833 0.5657376 0.1879231 0.05489595
[5,] 0.09223575 0.10238514 0.5670663 0.1561227 0.08219005
[6,] 0.09228184 0.10466584 0.6366615 0.1152634 0.05112739
```

```
R> apply(P, 2, mean)
```

```
[1] 0.07111111 0.09333333 0.63666666 0.14333334 0.05555556
```

This can be computed much more simply using the `fitted` function, with the `outcome` argument set to `FALSE` so that the probabilities for all the alternatives (and not only the chosen one) is returned.

```
R> apply(fitted(mc, outcome = FALSE), 2, mean)
```

```
      hp      ec      er      gc      gr
0.05555556 0.07111111 0.09333333 0.63666667 0.14333333
```

7. The California Energy Commission (CEC) is considering whether to offer rebates on heat pumps. The CEC wants to predict the effect of the rebates on the heating system choices of customers in California. The rebates will be set at 10% of the installation cost. Using the estimated coefficients from the model in exercise 6, calculate new probabilities and predicted shares using the new installation cost of heat pump. How much do the rebates raise the share of houses with heat pumps?

```
R> Hn <- H
R> Hn[Hn$alt == "hp", "ic"] <- 0.9 * Hn[Hn$alt == "hp", "ic"]
R> apply(predict(mc, newdata = Hn), 2, mean)
```

```
      hp      ec      er      gc      gr
0.06446230 0.07045486 0.09247026 0.63064443 0.14196814
```

We estimate the model with the actual costs. Then we change the costs and calculate probabilities with the new costs. The average probability is the predicted share for an alternative. At the original costs, the heat pump share is 0.0555 (ie, about 5.5%) This share is predicted to rise to 0.0645 (about 6.5%) when rebates are given.

8. Suppose a new technology is developed that provides more efficient central heating. The new technology costs \$200 more than the central electric system. However, it saves 25% of the electricity, such that its operating costs are 75% of the operating costs of `ec`. We want to predict the potential market penetration of this technology. Note that there are now six alternatives: the original five alternatives plus this new one. Calculate the probability and predict the market share (i.e., the average probability) for all six

alternatives, using the model that is estimated on the original five alternatives. (Be sure to use the original installation cost for heat pumps, rather than the reduced cost in exercise 7.) What is the predicted market share for the new technology? From which of the original five systems does the new technology draw the most customers?

```
R> X <- model.matrix(mc)
R> Xn <- X[alt == "ec",]
R> Xn[, "ic"] <- Xn[, "ic"] + 200
R> Xn[, "oc"] <- Xn[, "oc"] * 0.75
R> unchid <- unique(index(H)$chid)
R> rownames(Xn) <- paste(unchid, 'new', sep = ".")
R> chidb <- c(chid, unchid)
R> X <- rbind(X, Xn)
R> X <- X[order(chidb), ]
R> eXb <- as.numeric(exp(X %*% coef(mc)))
R> SeXb <- as.numeric(tapply(eXb, sort(chidb), sum))
R> P <- eXb / SeXb[sort(chidb)]
R> P <- matrix(P, ncol = 6, byrow = TRUE)
R> apply(P, 2, mean)
```

```
[1] 0.06311578 0.08347713 0.57145108 0.12855080 0.04977350 0.10363170
```

The new technology captures a market share of 0.1036. That is, it gets slightly more than ten percent of the market.

It draws the same percent (about 10%) from each system. This means that it draws the most in absolute terms from the most popular system, gas central. For example, gas central drops from 0.637 to 0.571; this is an absolute drop of $0.637 - 0.571 = 0.065$ and a percent drop of $0.065/0.637$ about 10%. Of the 10.36% market share that is attained by the new technology, 6.5% of it comes from gas central. The other systems drop by about the same percent, which is less in absolute terms.

The same percent drop for all systems is a consequence of the IIA property of logit. To me, this property seems unreasonable in this application. The new technology is a type of electric system. It seems reasonable that it would draw more from other electric systems than from gas systems. Models like nested logit, probit, and mixed logit allow more flexible, and in this case, more realistic substitution patterns.

2. Nested logit model

The data set **HC** from **mlogit** contains data in R format on the choice of heating and central cooling system for 250 single-family, newly built houses in California.

The alternatives are:

1. Gas central heat with cooling (**gcc**)
2. Electric central resistance heat with cooling (**ecc**)
3. Electric room resistance heat with cooling (**erc**)

4. Electric heat pump, which provides cooling also (`hpc`)
5. Gas central heat without cooling (`gc`)
6. Electric central resistance heat without cooling (`ec`)
7. Electric room resistance heat without cooling (`er`)

Heat pumps necessarily provide both heating and cooling such that heat pump without cooling is not an alternative.

The variables are:

- `depvar` gives the name of the chosen alternative,
- `ich.alt` are the installation cost for the heating portion of the system,
- `icca` is the installation cost for cooling
- `och.alt` are the operating cost for the heating portion of the system
- `occa` is the operating cost for cooling
- `income` is the annual income of the household

Note that the full installation cost of alternative `gcc` is `ich.gcc+icca`, and similarly for the operating cost and for the other alternatives with cooling.

1. Run a nested logit model on the data for two nests and one log-sum coefficient that applies to both nests. Note that the model is specified to have the cooling alternatives (`gcc`, `ecc`, `erc`, `hpc`) in one nest and the non-cooling alternatives (`gc`, `ec`, `er`) in another nest.

```
R> library("mlogit")
R> data("HC", package = "mlogit")
R> HC <- mlogit.data(HC, varying = c(2:8, 10:16), choice = "depvar", shape = "wide")
R> cooling.modes <- index(HC)$alt %in% c('gcc', 'ecc', 'erc', 'hpc')
R> room.modes <- index(HC)$alt %in% c('erc', 'er')
R> # installation / operating costs for cooling are constants,
R> # only relevant for mixed systems
R> HC$icca[!cooling.modes] <- 0
R> HC$occa[!cooling.modes] <- 0
R> # create income variables for two sets cooling and rooms
R> HC$inc.cooling <- HC$inc.room <- 0
R> HC$inc.cooling[cooling.modes] <- HC$income[cooling.modes]
R> HC$inc.room[room.modes] <- HC$income[room.modes]
R> # create an intercet for cooling modes
R> HC$int.cooling <- as.numeric(cooling.modes)
R> # estimate the model with only one nest elasticity
R> nl <- mlogit(depvar ~ ich + och +icca + occa + inc.room + inc.cooling + int.cooling | 0, HC,
+             nests = list(cooling = c('gcc', 'ecc', 'erc', 'hpc'),
+             other = c('gc', 'ec', 'er')), un.nest.el = TRUE)
R> summary(nl)
```

Call:

```
mlogit(formula = depvar ~ ich + och + icca + occa + inc.room +
        inc.cooling + int.cooling | 0, data = HC, nests = list(cooling = c("gcc",
        "ecc", "erc", "hpc"), other = c("gc", "ec", "er")), un.nest.el = TRUE)
```

Frequencies of alternatives:

```
      ec  ecc  er  erc  gc  gcc  hpc
0.004 0.016 0.032 0.004 0.096 0.744 0.104
```

bfgs method

11 iterations, 0h:0m:1s

$g'(-H)^{-1}g = 7.26E-06$

successive function values within tolerance limits

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
ich	-0.554878	0.144205	-3.8478	0.0001192 ***
och	-0.857886	0.255313	-3.3601	0.0007791 ***
icca	-0.225079	0.144423	-1.5585	0.1191212
occa	-1.089458	1.219821	-0.8931	0.3717882
inc.room	-0.378971	0.099631	-3.8038	0.0001425 ***
inc.cooling	0.249575	0.059213	4.2149	2.499e-05 ***
int.cooling	-6.000415	5.562423	-1.0787	0.2807030
iv	0.585922	0.179708	3.2604	0.0011125 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -178.12

(a) The estimated log-sum coefficient is -0.59 . What does this estimate tell you about the degree of correlation in unobserved factors over alternatives within each nest?

The correlation is approximately $1 - 0.59 = 0.41$. It's a moderate correlation.

(b) Test the hypothesis that the log-sum coefficient is 1.0 (the value that it takes for a standard logit model.) Can the hypothesis that the true model is standard logit be rejected?

We can use a t-test of the hypothesis that the log-sum coefficient equal to 1. The t-statistic is :

```
R> (coef(nl)['iv']-1)/sqrt(vcov(nl)['iv', 'iv'])
```

```
      iv
-2.304171
```

The critical value of t for 95% confidence is 1.96. So we can reject the hypothesis at 95% confidence.

We can also use a likelihood ratio test because the multinomial logit is a special case of the nested model.

```
R> # First estimate the multinomial logit model
```

```
R> ml <- update(nl, nests = NULL)
R> lrtest(nl, ml)
```

Likelihood ratio test

```
Model 1: depvar ~ ich + och + icca + occa + inc.room + inc.cooling + int.cooling |
0
```

```
Model 2: depvar ~ ich + och + icca + occa + inc.room + inc.cooling + int.cooling |
0
```

```
#Df  LogLik Df  Chisq Pr(>Chisq)
1    8 -178.12
2    7 -180.29 -1 4.3234  0.03759 *
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that the hypothesis is rejected at 95% confidence, but not at 99% confidence.

2. Re-estimate the model with the room alternatives in one nest and the central alternatives in another nest. (Note that a heat pump is a central system.)

```
R> nl2 <- update(nl, nests = list(central = c('ec', 'ecc', 'gc', 'gcc', 'hpc'),
+                               room = c('er', 'erc')))
R> summary(nl2)
```

Call:

```
mlogit(formula = depvar ~ ich + och + icca + occa + inc.room +
  inc.cooling + int.cooling | 0, data = HC, nests = list(central = c("ec",
  "ecc", "gc", "gcc", "hpc"), room = c("er", "erc")), un.nest.el = TRUE)
```

Frequencies of alternatives:

```
   ec  ecc  er  erc  gc  gcc  hpc
0.004 0.016 0.032 0.004 0.096 0.744 0.104
```

bfgs method

10 iterations, 0h:0m:1s

$g'(-H)^{-1}g = 5.87E-07$

gradient close to zero

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
ich	-1.13818	0.54216	-2.0993	0.03579 *
och	-1.82532	0.93228	-1.9579	0.05024 .
icca	-0.33746	0.26934	-1.2529	0.21024
occa	-2.06328	1.89726	-1.0875	0.27681
inc.room	-0.75722	0.34292	-2.2081	0.02723 *
inc.cooling	0.41689	0.20742	2.0099	0.04444 *

```
int.cooling -13.82487    7.94031 -1.7411  0.08167 .
iv          1.36201    0.65393  2.0828  0.03727 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Log-Likelihood: -180.02
```

(a) What does the estimate imply about the substitution patterns across alternatives? Do you think the estimate is plausible?

The log-sum coefficient is over 1. This implies that there is more substitution across nests than within nests. I don't think this is very reasonable, but people can differ on their concepts of what's reasonable.

(b) Is the log-sum coefficient significantly different from 1?

The t-statistic is :

```
R> (coef(nl2)['iv']-1)/sqrt(vcov(nl2)['iv', 'iv'])
```

```
      iv
0.5535849
```

```
R> lrtest(nl2, ml)
```

Likelihood ratio test

```
Model 1: depvar ~ ich + och + icca + occa + inc.room + inc.cooling + int.cooling |
0
Model 2: depvar ~ ich + och + icca + occa + inc.room + inc.cooling + int.cooling |
0
#Df  LogLik Df  Chisq Pr(>Chisq)
1    8 -180.02
2    7 -180.29 -1  0.5268    0.468
```

We cannot reject the hypothesis at standard confidence levels.

(c) How does the value of the log-likelihood function compare for this model relative to the model in exercise 1, where the cooling alternatives are in one nest and the heating alternatives in the other nest.

```
R> logLik(nl)
```

```
'log Lik.' -178.1247 (df=8)
```

```
R> logLik(nl2)
```

```
'log Lik.' -180.0231 (df=8)
```

The $\ln L$ is worse (more negative.) All in all, this seems like a less appropriate nesting structure.

3. Rerun the model that has the cooling alternatives in one nest and the non-cooling alternatives in the other nest (like for exercise 1), with a separate log-sum coefficient for each nest.

```
R> n13 <- update(n1, un.nest.el = FALSE)
```

(a) Which nest is estimated to have the higher correlation in unobserved factors? Can you think of a real-world reason for this nest to have a higher correlation?

The correlation in the cooling nest is around $1 - 0.60 = 0.4$ and that for the non-cooling nest is around $1 - 0.45 = 0.55$. So the correlation is higher in the non-cooling nest. Perhaps more variation in comfort when there is no cooling. This variation in comfort is the same for all the non-cooling alternatives.

(b) Are the two log-sum coefficients significantly different from each other? That is, can you reject the hypothesis that the model in exercise 1 is the true model?

We can use a likelihood ratio tests with models `n1` and `n13`.

```
R> lrtest(n1, n13)
```

Likelihood ratio test

```
Model 1: depvar ~ ich + och + icca + occa + inc.room + inc.cooling + int.cooling |
0
Model 2: depvar ~ ich + och + icca + occa + inc.room + inc.cooling + int.cooling |
0
#Df  LogLik Df  Chisq Pr(>Chisq)
1    8 -178.12
2    9 -178.04  1 0.1758      0.675
```

The restricted model is the one from exercise 1 that has one log-sum coefficient. The unrestricted model is the one we just estimated. The test statistics is 0.6299. The critical value of chi-squared with 1 degree of freedom is 3.8 at the 95% confidence level. We therefore cannot reject the hypothesis that the two nests have the same log-sum coefficient.

4. Rewrite the code to allow three nests. For simplicity, estimate only one log-sum coefficient which is applied to all three nests. Estimate a model with alternatives `gcc`, `ecc` and `erc` in a nest, `hpc` in a nest alone, and alternatives `gc`, `ec` and `er` in a nest. Does this model seem better or worse than the model in exercise 1, which puts alternative `hpc` in the same nest as alternatives `gcc`, `ecc` and `erc`?

```
R> n14 <- update(n1, nests=list(n1 = c('gcc', 'ecc', 'erc'), n2 = c('hpc'),
+                               n3 = c('gc', 'ec', 'er')))
R> summary(n14)
```

Call:

```
mlogit(formula = depvar ~ ich + och + icca + occa + inc.room +
inc.cooling + int.cooling | 0, data = HC, nests = list(n1 = c("gcc",
"ecc", "erc"), n2 = c("hpc"), n3 = c("gc", "ec", "er")),
```

```

un.nest.el = TRUE)

Frequencies of alternatives:
   ec  ecc  er  erc  gc  gcc  hpc
0.004 0.016 0.032 0.004 0.096 0.744 0.104

bfgs method
8 iterations, 0h:0m:1s
g'(-H)^-1g = 3.71E-08
gradient close to zero

Coefficients :
      Estimate Std. Error t-value Pr(>|t|)
ich          -0.838394   0.100546 -8.3384 < 2.2e-16 ***
och          -1.331598   0.252069 -5.2827 1.273e-07 ***
icca         -0.256131   0.145564 -1.7596  0.07848 .
occa         -1.405656   1.207281 -1.1643  0.24430
inc.room     -0.571352   0.077950 -7.3297 2.307e-13 ***
inc.cooling   0.311355   0.056357  5.5247 3.301e-08 ***
int.cooling  -10.413384   5.612445 -1.8554  0.06354 .
iv            0.956544   0.180722  5.2929 1.204e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -180.26

```

The $\ln L$ for this model is -180.26 , which is lower (more negative) than for the model with two nests, which got -178.12 .

3. Mixed logit model

A sample of residential electricity customers were asked a series of choice experiments. In each experiment, four hypothetical electricity suppliers were described. The person was asked which of the four suppliers he/she would choose. As many as 12 experiments were presented to each person. Some people stopped before answering all 12. There are 361 people in the sample, and a total of 4308 experiments. In the experiments, the characteristics of each supplier were stated. The price of the supplier was either :

1. a fixed price at a stated cents per kWh, with the price varying over suppliers and experiments.
2. a time-of-day (TOD) rate under which the price is 11 cents per kWh from 8am to 8pm and 5 cents per kWh from 8pm to 8am. These TOD prices did not vary over suppliers or experiments: whenever the supplier was said to offer TOD, the prices were stated as above.

3. a seasonal rate under which the price is 10 cents per kWh in the summer, 8 cents per kWh in the winter, and 6 cents per kWh in the spring and fall. Like TOD rates, these prices did not vary. Note that the price is for the electricity only, not transmission and distribution, which is supplied by the local regulated utility.

The length of contract that the supplier offered was also stated, in years (such as 1 year or 5 years.) During this contract period, the supplier guaranteed the prices and the buyer would have to pay a penalty if he/she switched to another supplier. The supplier could offer no contract in which case either side could stop the agreement at any time. This is recorded as a contract length of 0.

Some suppliers were also described as being a local company or a “well-known” company. If the supplier was not local or well-known, then nothing was said about them in this regard ¹.

1. Run a mixed logit model without intercepts and a normal distribution for the 6 parameters of the model, using 100 draws, halton sequences and taking into account the panel data structure.

```
R> library("mlogit")
R> data("Electricity", package = "mlogit")
R> Electr <- mlogit.data(Electricity, id="id", choice="choice",
+                       varying=3:26, shape="wide", sep="")

R> Elec.mxl <- mlogit(choice~pf+cl+loc+wk+tod+seas|0, Electr,
+                   rpar=c(pf='n', cl='n', loc='n', wk='n', tod='n', seas='n'),
+                   R=100, halton=NA, print.level=0, panel=TRUE)
```

```
R> summary(Elec.mxl)
```

Call:

```
mlogit(formula = choice ~ pf + cl + loc + wk + tod + seas | 0,
       data = Electr, rpar = c(pf = "n", cl = "n", loc = "n", wk = "n",
       tod = "n", seas = "n"), R = 100, halton = NA, panel = TRUE,
       print.level = 0)
```

Frequencies of alternatives:

	1	2	3	4
	0.22702	0.26393	0.23816	0.27089

bfgs method

21 iterations, 0h:2m:19s

$g'(-H)^{-1}g = 3.3E-07$

¹These are the data used in [Revelt and Train \(2000\)](#) and [Hubert and Train \(2001\)](#).

gradient close to zero

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
pf	-0.973384	0.034324	-28.359	< 2.2e-16 ***
cl	-0.205557	0.013323	-15.428	< 2.2e-16 ***
loc	2.075733	0.080430	25.808	< 2.2e-16 ***
wk	1.475650	0.065168	22.644	< 2.2e-16 ***
tod	-9.052542	0.287219	-31.518	< 2.2e-16 ***
seas	-9.103772	0.289043	-31.496	< 2.2e-16 ***
sd.pf	0.219945	0.010840	20.291	< 2.2e-16 ***
sd.cl	0.378304	0.018489	20.461	< 2.2e-16 ***
sd.loc	1.482980	0.081305	18.240	< 2.2e-16 ***
sd.wk	1.000061	0.074182	13.481	< 2.2e-16 ***
sd.tod	2.289489	0.110731	20.676	< 2.2e-16 ***
sd.seas	1.180883	0.109007	10.833	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -3952.5

random coefficients

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
pf	-Inf	-1.1217350	-0.9733844	-0.9733844	-0.82503376	Inf
cl	-Inf	-0.4607190	-0.2055565	-0.2055565	0.04960589	Inf
loc	-Inf	1.0754783	2.0757333	2.0757333	3.07598832	Inf
wk	-Inf	0.8011189	1.4756497	1.4756497	2.15018054	Inf
tod	-Inf	-10.5967791	-9.0525423	-9.0525423	-7.50830550	Inf
seas	-Inf	-9.9002649	-9.1037717	-9.1037717	-8.30727842	Inf

2. (a) Using the estimated mean coefficients, determine the amount that a customer with average coefficients for price and length is willing to pay for an extra year of contract length.

```
R> coef(Elec.mxl)['cl']/coef(Elec.mxl)['pf']
```

```
      cl
0.2111772
```

The mean coefficient of length is -0.20. The consumer with this average coefficient dislikes having a longer contract. So this person is willing to pay to reduce the length of the contract. The mean price coefficient is -0.97. A customer with these coefficients is willing to pay $0.20/0.97=0.21$, or one-fifth a cent per kWh extra to have a contract that is one year shorter.

(b) Determine the share of the population who are estimated to dislike long term contracts (ie have a negative coefficient for the length.)

```
R> pnorm(-coef(Elec.mx1)['cl']/coef(Elec.mx1)['sd.cl'])
```

```
      cl
0.70656
```

The coefficient of length is normally distributed with mean -0.20 and standard deviation 0.35. The share of people with coefficients below zero is the cumulative probability of a standardized normal deviate evaluated at $0.20/0.35=0.57$. Looking 0.57 up in a table of the standard normal distribution, we find that the share below 0.57 is 0.72. About seventy percent of the population are estimated to dislike long-term contracts.

3. The price coefficient is assumed to be normally distributed in these runs. This assumption means that some people are assumed to have positive price coefficients, since the normal distribution has support on both sides of zero. Using your estimates from exercise 1, determine the share of customers with positive price coefficients. As you can see, this is pretty small share and can probably be ignored. However, in some situations, a normal distribution for the price coefficient will give a fairly large share with the wrong sign. Revise the program to make the price coefficient fixed rather than random. A fixed price coefficient also makes it easier to calculate the distribution of willingness to pay (wtp) for each non-price attribute. If the price coefficient is fixed, the distribution of wtp for an attribute has the same distribution as the attribute's coefficient, simply scaled by the price coefficient. However, when the price coefficient is random, the distribution of wtp is the ratio of two distributions, which is harder to work with.

```
R> pnorm(-coef(Elec.mx1)['pf']/coef(Elec.mx1)['sd.pf'])
```

```
      pf
0.9999952
```

The price coefficient is distributed normal with mean -0.97 and standard deviation 0.20. The cumulative standard normal distribution evaluated at $0.97/0.20=4.85$ is more than 0.999, which means that more than 99.9% of the population are estimated to have negative price coefficients. Essentially no one is estimated to have a positive price coefficient.

```
R> Elec.mx12 <- mlogit(choice~pf+cl+loc+wk+tod+seas|0, Electr,
+                    rpar=c(cl='n', loc='n', wk='n', tod='n', seas='n'),
+                    R=100, halton=NA, print.level=0, panel=TRUE)
```

```
R> summary(Elec.mx12)
```

Call:

```
mlogit(formula = choice ~ pf + cl + loc + wk + tod + seas | 0,
      data = Electr, rpar = c(cl = "n", loc = "n", wk = "n", tod = "n",
      seas = "n"), R = 100, halton = NA, panel = TRUE, print.level = 0)
```

Frequencies of alternatives:

```
      1      2      3      4
0.22702 0.26393 0.23816 0.27089
```

bfgs method

```
22 iterations, 0h:2m:21s
g'(-H)^-1g = 3.77E-07
gradient close to zero
```

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
pf	-0.879904	0.032759	-26.860	< 2.2e-16 ***
cl	-0.217060	0.013673	-15.875	< 2.2e-16 ***
loc	2.092292	0.081067	25.809	< 2.2e-16 ***
wk	1.490894	0.065230	22.856	< 2.2e-16 ***
tod	-8.581857	0.282912	-30.334	< 2.2e-16 ***
seas	-8.583296	0.280347	-30.617	< 2.2e-16 ***
sd.cl	0.373478	0.018018	20.728	< 2.2e-16 ***
sd.loc	1.558858	0.087696	17.776	< 2.2e-16 ***
sd.wk	1.050811	0.078023	13.468	< 2.2e-16 ***
sd.tod	2.694667	0.120799	22.307	< 2.2e-16 ***
sd.seas	1.950727	0.104766	18.620	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -3961.7

random coefficients

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
cl	-Inf	-0.4689672	-0.2170603	-0.2170603	0.0348465	Inf
loc	-Inf	1.0408581	2.0922916	2.0922916	3.1437251	Inf
wk	-Inf	0.7821322	1.4908937	1.4908937	2.1996553	Inf
tod	-Inf	-10.3993820	-8.5818566	-8.5818566	-6.7643312	Inf
seas	-Inf	-9.8990409	-8.5832956	-8.5832956	-7.2675502	Inf

4. You think that everyone must like using a known company rather than an unknown one, and yet the normal distribution implies that some people dislike using a known company. Revise the program to give the coefficient of *wk* a uniform distribution (do this with the price coefficient fixed).

```
R> Elec.mx13 <- update(Elec.mx1, rpar=c(cl='n', loc='n', wk='u', tod='n', seas='n'))
```

The price coefficient is uniformly distributed with parameters 1.541 and 1.585.

```
R> summary(Elec.mx13)
```

Call:

```
mlogit(formula = choice ~ pf + cl + loc + wk + tod + seas | 0,
        data = Electr, rpar = c(cl = "n", loc = "n", wk = "u", tod = "n",
                               seas = "n"), R = 100, halton = NA, panel = TRUE, print.level = 0)
```

Frequencies of alternatives:

	1	2	3	4
	0.22702	0.26393	0.23816	0.27089

```
bfgs method
22 iterations, 0h:2m:27s
g'(-H)^-1g = 3.71E-07
gradient close to zero
```

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
pf	-0.882232	0.032818	-26.883	< 2.2e-16 ***
cl	-0.217127	0.013776	-15.761	< 2.2e-16 ***
loc	2.099319	0.081056	25.900	< 2.2e-16 ***
wk	1.509410	0.065496	23.046	< 2.2e-16 ***
tod	-8.607002	0.282983	-30.415	< 2.2e-16 ***
seas	-8.602408	0.280671	-30.649	< 2.2e-16 ***
sd.cl	0.381070	0.018150	20.996	< 2.2e-16 ***
sd.loc	1.593850	0.087802	18.153	< 2.2e-16 ***
sd.wk	1.786377	0.125764	14.204	< 2.2e-16 ***
sd.tod	2.719078	0.119357	22.781	< 2.2e-16 ***
sd.seas	1.945377	0.103686	18.762	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -3956.7

random coefficients

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
cl	-Inf	-0.4741552	-0.2171273	-0.2171273	0.03990058	Inf
loc	-Inf	1.0242834	2.0993191	2.0993191	3.17435473	Inf
wk	-0.2769665	0.6162218	1.5094101	1.5094101	2.40259840	3.295787
tod	-Inf	-10.4409924	-8.6070022	-8.6070022	-6.77301190	Inf
seas	-Inf	-9.9145449	-8.6024084	-8.6024084	-7.29027193	Inf

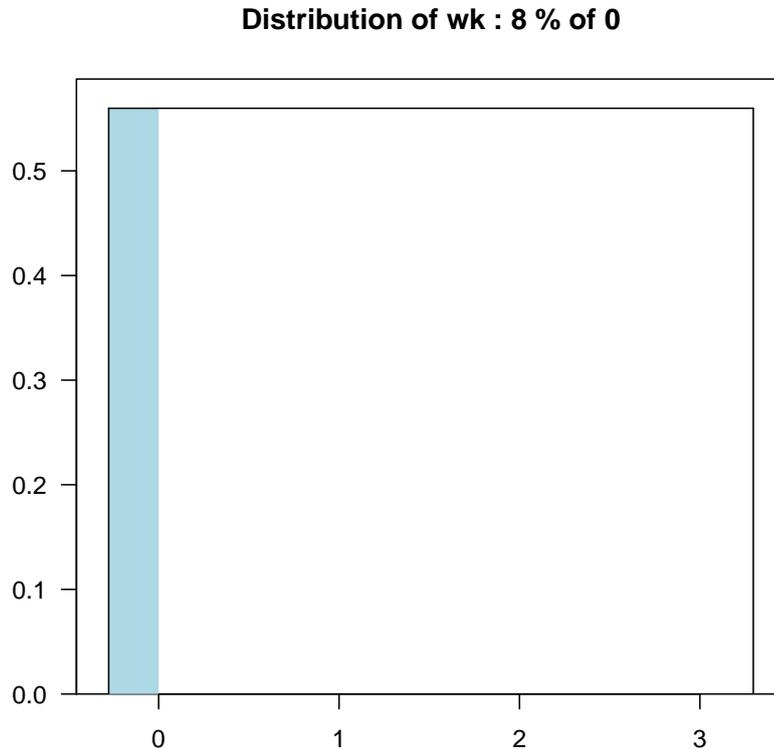
```
R> rpar(Elec.mx13, 'wk')
```

uniform distribution with parameters 1.509 (center) and 1.786 (span)

```
R> summary(rpar(Elec.mx13, 'wk'))
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	-0.2769665	0.6162218	1.5094101	1.5094101	2.4025984	3.2957867

```
R> plot(rpar(Elec.mx13, 'wk'))
```



The upper bound is 3.13. The estimated price coefficient is -0.88 and so the willingness to pay for a known provided ranges uniformly from -0.05 to 3.55 cents per kWh.

5. Rerun the model with a fixed coefficient for price and lognormal distributions for the coefficients of TOD and seasonal (since their coefficient should be negative for all people.) To do this, you need to reverse the sign of the TOD and seasonal variables, since the lognormal is always positive and you want these coefficients to be always negative.

A lognormal is specified as $\exp(b + se)$ where e is a standard normal deviate. The parameters of the lognormal are b and s . The mean of the lognormal is $\exp(b + 0.5s^2)$ and the standard deviation is the mean times $\sqrt{(\exp(s^2)) - 1}$.

```
R> Electr <- mlogit.data(Electricity, id="id", choice="choice",
+                       varying=3:26, shape="wide", sep="",
+                       opposite=c('tod', 'seas'))
R> Elec.mx14 <- mlogit(choice~pf+cl+loc+wk+tod+seas|0, Electr,
+                   rpar=c(cl='n', loc='n', wk='u', tod='ln', seas='ln'),
+                   R=100, halton=NA, print.level=0, panel=TRUE)

R> summary(Elec.mx14)
```

Call:

```
mlogit(formula = choice ~ pf + cl + loc + wk + tod + seas | 0,
       data = Electr, rpar = c(cl = "n", loc = "n", wk = "u", tod = "ln",
                               seas = "ln"), R = 100, halton = NA, panel = TRUE, print.level = 0)
```

Frequencies of alternatives:

```
      1      2      3      4
0.22702 0.26393 0.23816 0.27089
```

bfgs method

20 iterations, 0h:2m:14s

$g'(-H)^{-1}g = 2.97E-07$

gradient close to zero

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)
pf	-0.868987	0.032350	-26.863	< 2.2e-16 ***
cl	-0.211333	0.013569	-15.575	< 2.2e-16 ***
loc	2.023888	0.080102	25.266	< 2.2e-16 ***
wk	1.479124	0.064957	22.771	< 2.2e-16 ***
tod	2.112381	0.033769	62.554	< 2.2e-16 ***
seas	2.124207	0.033342	63.710	< 2.2e-16 ***
sd.cl	0.373120	0.017710	21.068	< 2.2e-16 ***
sd.loc	1.548510	0.086400	17.922	< 2.2e-16 ***
sd.wk	1.521792	0.119994	12.682	< 2.2e-16 ***
sd.tod	0.367076	0.019997	18.357	< 2.2e-16 ***
sd.seas	0.275350	0.016875	16.317	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

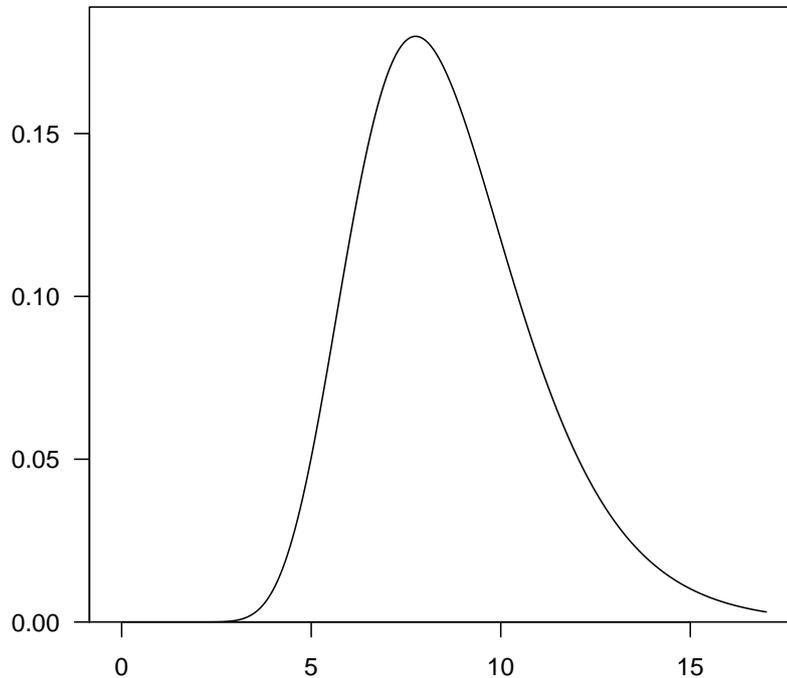
Log-Likelihood: -3976.5

random coefficients

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
cl	-Inf	-0.4629985	-0.2113327	-0.2113327	0.04033302	Inf
loc	-Inf	0.9794338	2.0238880	2.0238880	3.06834218	Inf
wk	-0.04266822	0.7182277	1.4791236	1.4791236	2.24001959	3.000916
tod	0.00000000	6.4545926	8.2679047	8.8441271	10.59063726	Inf
seas	0.00000000	6.9482273	8.3662614	8.6895038	10.07369591	Inf

```
R> plot(rpar(Elec.mx14, 'seas'))
```

Distribution of seas



6. Rerun the same model as previously, but allowing now the correlation between random parameters. Compute the correlation matrix of the random parameters. Test the hypothesis that the random parameters are uncorrelated.

```
R> Elec.mx15 <- update(Elec.mx14, correlation = TRUE)
```

```
R> summary(Elec.mx15)
```

Call:

```
mlogit(formula = choice ~ pf + cl + loc + wk + tod + seas | 0,
        data = Electr, rpar = c(cl = "n", loc = "n", wk = "u", tod = "ln",
                                seas = "ln"), R = 100, correlation = TRUE, halton = NA,
        panel = TRUE, print.level = 0)
```

Frequencies of alternatives:

1	2	3	4
0.22702	0.26393	0.23816	0.27089

bfgs method

28 iterations, 0h:3m:27s

$g'(-H)^{-1}g = 5.54E-07$

gradient close to zero

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)	
pf	-0.9177040	0.0340200	-26.9754	< 2.2e-16	***
cl	-0.2158517	0.0138412	-15.5948	< 2.2e-16	***
loc	2.3925710	0.0869030	27.5315	< 2.2e-16	***
wk	1.7475319	0.0712088	24.5409	< 2.2e-16	***
tod	2.1554624	0.0337210	63.9205	< 2.2e-16	***
seas	2.1695481	0.0334582	64.8436	< 2.2e-16	***
cl.cl	0.3962523	0.0187077	21.1813	< 2.2e-16	***
cl.loc	0.6174971	0.0924280	6.6808	2.376e-11	***
cl.wk	0.1952382	0.0731907	2.6675	0.007641	**
cl.tod	0.0019823	0.0104183	0.1903	0.849095	
cl.seas	0.0259964	0.0098345	2.6434	0.008208	**
loc.loc	-2.0717181	0.1063248	-19.4848	< 2.2e-16	***
loc.wk	-1.2366645	0.0866099	-14.2786	< 2.2e-16	***
loc.tod	0.0625084	0.0119610	5.2260	1.732e-07	***
loc.seas	-0.0012253	0.0098998	-0.1238	0.901494	
wk.wk	0.6431903	0.0742356	8.6642	< 2.2e-16	***
wk.tod	0.1606723	0.0138056	11.6382	< 2.2e-16	***
wk.seas	0.1413816	0.0128752	10.9809	< 2.2e-16	***
tod.tod	0.3758556	0.0209478	17.9425	< 2.2e-16	***
tod.seas	0.0899901	0.0109770	8.1980	2.220e-16	***
seas.seas	0.2112446	0.0141904	14.8865	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -3851.4

random coefficients

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
cl	-Inf	-0.4831199	-0.2158517	-0.2158517	0.0514164	Inf
loc	-Inf	0.9344685	2.3925710	2.3925710	3.8506735	Inf
wk	0.3399982	1.0437651	1.7475319	1.7475319	2.4512987	3.155065
tod	0.0000000	6.5309411	8.6318806	9.4023489	11.4086716	Inf
seas	0.0000000	7.2923557	8.7543271	9.0815274	10.5093945	Inf

R> cor.mlogit(Elec.mx15)

	cl	loc	wk	tod	seas
cl	1.00000000	0.28564213	0.13870943	0.00479382	0.09596179
loc	0.28564213	1.00000000	0.88161933	-0.14349636	0.03174548
wk	0.13870943	0.88161933	1.00000000	0.04540614	0.25576901
tod	0.00479382	-0.14349636	0.04540614	1.00000000	0.50449142
seas	0.09596179	0.03174548	0.25576901	0.50449142	1.00000000

R> lrtest(Elec.mx15, Elec.mx14)

Likelihood ratio test

```

Model 1: choice ~ pf + cl + loc + wk + tod + seas | 0
Model 2: choice ~ pf + cl + loc + wk + tod + seas | 0
#Df LogLik Df Chisq Pr(>Chisq)
1 21 -3851.4
2 11 -3976.5 -10 250.18 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
R> waldtest(Elec.mxl5, correlation = FALSE)
```

Wald test

```

data: uncorrelated random effects
chisq = 466.4757, df = 10, p-value < 2.2e-16

```

```
R> scoretest(Elec.mxl4, correlation = TRUE)
```

score test

```

data: correlation = TRUE
chisq = 2794.891, df = 10, p-value < 2.2e-16
alternative hypothesis: uncorrelated random effects

```

The three tests clearly reject the hypothesis that the random parameters are uncorrelated.

4. Multinomial probit

We have data on the mode choice of 453 commuters. Four modes are available: (1) bus, (2) car alone, (3) carpool, and (4) rail. We have data for each commuter on the cost and time on each mode and the chosen mode.

`mlogit` estimates the multinomial probit model if the `probit` argument is `TRUE` using the GHK procedure. This program estimates the full covariance matrix subject to normalization.

More precisely, utility differences are computed respective to the reference level of the response (by default the bus alternative) and the 3×3 matrix of covariance is estimated. As the scale of utility is unobserved, the first element of the matrix is further set to 1. The Choleski factor of the covariance is :

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \theta_{32} & \theta_{33} & 0 \\ \theta_{42} & \theta_{43} & \theta_{44} \end{pmatrix}$$

that is, five covariance parameters are estimated. The covariance matrix of the utility differences is then $\Omega = LL^\top$. By working in a Choleski factor that has this form, the

normalization constraints are automatically imposed and the covariance matrix is guaranteed to be positive semi-definite (with the covariance matrix for any error differences being positive-definite).

2. Estimate a model where mode choice is explained by the time and the cost of an alternative, using 100 draws and set the seed to 20. Calculate the covariance matrix that is implied by the estimates of the Choleski factor. What, if anything, can you say about the degree of heteroskedasticity and correlation? For comparison, what would the covariance matrix be if there were no heteroskedasticity or correlation (ie, iid errors for each alternative)? Can you tell whether the covariance is higher between car alone and carpool or between bus and rail?

```
R> data("Mode", package="mlogit")
R> Mo <- mlogit.data(Mode, choice='choice', shape='wide', varying=c(2:9))

R> p1 <- mlogit(choice~cost+time, Mo, seed = 20, R = 100, probit = TRUE)

R> summary(p1)
```

Call:

```
mlogit(formula = choice ~ cost + time, data = Mo, probit = TRUE,
       R = 100, seed = 20)
```

Frequencies of alternatives:

```
      bus      car carpool      rail
0.17881 0.48124 0.07064 0.26932
```

bfgs method

20 iterations, 0h:1m:22s

$g'(-H)^{-1}g = 7.71E-07$

gradient close to zero

Coefficients :

	Estimate	Std. Error	t-value	Pr(> t)	
car:(intercept)	1.8308660	0.2506434	7.3047	2.780e-13	***
carpool:(intercept)	-1.2816819	0.5677813	-2.2574	0.0239861	*
rail:(intercept)	0.3093510	0.1151701	2.6860	0.0072305	**
cost	-0.4134401	0.0731593	-5.6512	1.593e-08	***
time	-0.0466552	0.0068263	-6.8347	8.220e-12	***
car.carpool	0.2599724	0.3850288	0.6752	0.4995472	
car.rail	0.7364869	0.2145744	3.4323	0.0005985	***
carpool.carpool	1.3078947	0.3916729	3.3393	0.0008400	***
carpool.rail	-0.7981842	0.3463735	-2.3044	0.0212000	*
rail.rail	0.4301303	0.4874624	0.8824	0.3775677	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -347.92

```
McFadden R^2: 0.36012
Likelihood ratio test : chisq = 391.62 (p.value = < 2.22e-16)
```

The estimated Choleski factor L_1 is :

```
R> L1 <- matrix(0, 3, 3)
R> L1[!upper.tri(L1)] <- c(1, coef(p1)[6:10])
```

Multiplying L1 by its transpose gives Ω_1 :

```
R> L1 %**% t(L1)

      [,1] [,2] [,3]
[1,] 1.0000000 1.0000000 0.2599724
[2,] 1.0000000 1.542413 1.2232198
[3,] 0.2599724 1.223220 2.4152722
```

With iid errors, Ω_1 would be :

$$\begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}$$

I find it hard to tell anything from the estimated covariance terms.

3. Change the seed to 21 and rerun the model. (Even though the seed is just one digit different, the random draws are completely different.) See how much the estimates change. Does there seem to be a relation between the standard error of a parameter and the amount that its estimate changes with the new seed? (Of course, you are only getting two estimates, and so you are not estimating the true simulation variance very well. But the two estimates will probably give you an indication.)

```
R> p2 <- mlogit(choice~cost+time, Mo, seed = 21, R = 100, probit = TRUE)
```

```
R> coef(p2)
```

```
car:(intercept) carpool:(intercept) rail:(intercept) cost
1.87149948 -1.28893595 0.31455318 -0.43068703
time car.car car.carpool car.rail
-0.04752315 1.00000000 0.22888163 0.69781113
carpool.carpool carpool.rail rail.rail
1.33071717 -0.56802431 0.71060138
attr(,"fixed")
car:(intercept) carpool:(intercept) rail:(intercept) cost
FALSE FALSE FALSE FALSE
time car.car car.carpool car.rail
FALSE TRUE FALSE FALSE
carpool.carpool carpool.rail rail.rail
FALSE FALSE FALSE
```

The estimates seem to change more for parameters with larger standard error, though this is not uniformly the case by any means. One would expect larger samplign variance (which arises from a flatter $\ln L$ near the max) to translate into greater simulation variance (which raises when the location of the max changes with different draws).

4. Compute the probit shares (average probabilities) under user-specified parameters and data. How well do predicted shares match the actual share of commuters choosing each mode?

The actual shares are :

```
R> actShares <- with(Mo, tapply(choice, alt, mean))
```

The predicted shares are :

```
R> predShares <- apply(fitted(p1, outcome = FALSE), 2, mean)
R> predShares
```

```
      bus      car  carpool      rail
0.17596234 0.48239576 0.06923727 0.27209773
```

```
R> sum(predShares)
```

```
[1] 0.9996931
```

The correspondence is very close but not exact.

Note: Simulated ghk probabilities do not necessarily sum to one over alternatives. This summing-up error at the individual level tends to cancel out when the probabilities are averaged over the sample. The forecasted shares (ie, average probabilities) sum to 0.9991102, which is only slightly different from 1.

5. We want to examine the impact of a large tax on driving alone. Raise the cost of the car alone mode by 50% and forecast shares at these higher costs. Is the substitution proportional, as a logit model would predict? Which mode has the greatest percent increase in demand? What is the percent change in share for each mode?

```
R> Mo2 <- Mo
R> Mo2[Mo2$alt == 'car', 'cost'] <- Mo2[Mo2$alt == 'car', 'cost'] * 2
R> newShares <- apply(predict(p1, newdata = Mo2), 2, mean)
R> cbind(original = actShares, new = newShares, change = round((newShares - actShares) / actShares * 100))
```

```
      original      new change
bus      0.17880795 0.2517897    41
car      0.48123620 0.1689331   -65
carpool  0.07064018 0.1432554   103
rail     0.26931567 0.4356097    62
```

Substitution is not proportional. Carpool gets the largest percent increase.

References

- Hubert J, Train K (2001). "On the similarity of classical and Bayesian estimates of individual mean pathworths." *Marketing Letters*, **12**, 259–269.
- Revelt D, Train K (2000). "Customer-specific taste parameters and mixed logit'." Working Paper no. E00-274, Department of Economics, University of California, Berkeley.

Contents

Multinomial logit model	1
Nested logit model	10
Mixed logit model	16
Multinomial probit	26

Affiliation:

Yves Croissant
Faculté de Droit et d'Economie
Université de la Réunion
15, avenue René Cassin
BP 7151
F-97715 Saint-Denis Messag Cedex 9
Telephone: +33/262/938446
E-mail: yves.croissant@univ-reunion.fr