Package ‘RealVAMS’

February 19, 2015

Type Package
Title Multivariate VAM Fitting
Version 0.3-1
Date 2014-09-25
Author Andrew Karl, Jennifer Broatch, and Jennifer Green
Maintainer Andrew Karl <akarl@asu.edu>
Description The RealVAMS package fits a multivariate value-added model (VAM) (see Broatch and Lohr 2012) with normally distributed test scores and a binary outcome indicator. This material is based upon work supported by the National Science Foundation under grants DRL-1336027 and DRL-1336265.
License GPL-2
Depends R (>= 3.0.0), Matrix
Imports numDeriv, Rcpp (>= 0.10.6)
LazyData yes
ByteCompile yes
NeedsCompilation yes
LinkingTo Rcpp, RcppArmadillo
Repository CRAN
Date/Publication 2014-11-01 07:19:08

R topics documented:

RealVAMS-package .................................................. 2
example.outcome.data ............................................. 3
example.score.data .................................................. 4
RealVAMS ............................................................. 5
R_mstep2 ............................................................. 8
vp_cp ................................................................. 9

Index 10
RealVAMS-package

Multivariate VAM Fitting

Description

The RealVAMS package fits a multivariate value-added model (VAM) (see Broatch and Lohr 2012) with normally distributed test scores and a binary outcome indicator. This material is based upon work supported by the National Science Foundation under grants DRL-1336027 and DRL-1336265.

Details

- **Package**: RealVAMS
- **Type**: Package
- **Version**: 0.3-1
- **Date**: 2014-09-25
- **License**: GPL-2

Author(s)

Authors: Andrew Karl, Jennifer Broatch, and Jennifer Green

Maintainer: Andrew Karl <akarl@asu.edu>

References


Examples

```r
data(example.score.data)
data(example.outcome.data)
# The next line exists to show that the function can run and that the package
# installed correctly
res.test<-RealVAMS(example.score.data,example.outcome.data,max.PQL.it=1,max.iter.EM=2,
  var.parm.hessian=FALSE)
# The next line (not run automatically) provides a full example of the function
## Not run: res<RealVAMS(example.score.data,example.outcome.data)
```

---

**example.outcome.data**  **Simulated Data**

### Description

A simulated data set used to illustrate the functionality of the package. This data set represents binary outcome measurements on 625 students (with one missing).

### Usage

```r
data(example.outcome.data)
```

### Format

A data frame with 624 observations. The data set contains the following 2 variables.

- `r` a numeric vector composed of 0’s and 1’s representing a binary outcome measured on students.
- `student` a numeric vector

### Details

The data set may be reproduced with the following code.

```r
set.seed(0) library(MASS) years<-3 # teacher in each year teachers<-25 # students in each class
students<-25 alpha<-.5 etatruej<mvnorm(n=teachers*students,mu=c(0,0),Sigma=cbind(c(5,.2),c(.2,.1)))
etastrue<-eta.truem,1] eta.truat<- sample(rep(1:teachers,each=students),2)<sample(rep(1:teachers,each=students))
z3<-sample(rep(1:teachers,each=students)) cont_var1<-rmvnorm(students*teachers,0,5) cont_var2<-
  rmvnorm(students*teachers,0,5) cont_var3<rmvnorm(students*teachers,0,5) gam<rmvnorm(n=teachers*years,mu=c(0,0),Sigma=cbind(c(5,.6),c(.6,.6)))
eps1<rmvnorm(0,0,5) eps2<rmvnorm(0,0,5) eps3<rmvnorm(0,0,5)
gam1<gam[seq(1:teachers),1] gam2<-gam[seq((teachers+1),(2*teachers)),1] gam3<-gam[seq((2*teachers+1),(3*teachers)),1]
```

---

This code snippet demonstrates the creation of a simulated data set used to illustrate the functionality of the package. The data set represents binary outcome measurements on 625 students (with one missing).
Examples

data(example.score.data)
print(example.score.data[1,])

description

A simulated data set used to illustrate the functionality of the package. The data are simulated according to the VP model.

Usage

data(example.score.data)

Format

A data frame with 1874 observations on 625 students over 3 years, with 25 teachers in each year. The data set contains the following 5 variables.

- `y` a numeric vector representing the student score
- `student` a numeric vector
- `year` a numeric vector
- `teacher` a numeric vector
- `cont_var` a numeric vector representing a continuous covariate

Details

The data set may be reproduced with the following code.

```r
set.seed(0)
library(MASS)
# number of years: fixed at 3
# teacher in each year
teachers<-25
# students in each class
students<-25
alpha<-.5
eta.stu.j <- mvrnorm(n=teachers*students,mu=c(0,0),Sigma=cbind(c(5,.2),c(.2,.1)))
eta.stu <- eta.stu.j[,1]
eta.stu.r <- eta.stu.j[,2]
z1 <- rep(1:teachers,each=students)
z2 <- sample(rep(1:teachers,each=students))
z3 <- sample(rep(1:teachers,each=students))
cont_var1 <- rnorm(students*teachers,0,.5)
cont_var2 <- rnorm(students*teachers,0,.5)
gam <- mvrnorm(n=teachers*years,mu=c(0,0),Sigma=cbind(c(5,.6),c(.6,.6)))
eps1 <- rnorm(students*teachers,0,sqrt(5))
eps2 <- rnorm(students*teachers,0,sqrt(5))
eps3 <- rnorm(students*teachers,0,sqrt(5))
gam1 <- gam[seq(1,teachers),1]
gam2 <- gam[seq((teachers+1),(2*teachers)),1]
gam3 <- gam[seq((2*teachers+1),(3*teachers)),1]
y1 <- 50 + eta.stu + gam1[z1] + cont_var1 + eps1
y2 <- eta.stu + gam1[z1]*alpha + gam2[z2] + cont_var2 + eps2
y3 <- 100 + eta.stu + gam1[z1]*alpha + gam2[z2]*alpha + gam3[z3] + cont_var3 + eps3
r1 <- rbinom(students*teachers,1,pmnorm(.1 + eta.stu.r + gam1.r[z1] + gam2.r[z2] + gam3.r[z3]))
student<-1:(students*teachers)
year<-(rep(1:3,each=students*teachers))
student2<-as.data.frame(cbind(student,year,y))
vam_data2<-as.data.frame(cbind(student=student2$student,vam_data2$year),1)
vam_data2[]<-(vam_data2$year)
vam_data2.r<-as.data.frame(cbind(student,r=r1))
vam_data2.r<-vam_data2.r[-6,]
```
Examples

```r
data(example.score.data)
print(example.score.data[1,])
```

---

**RealVAMS**  
**Multivariate VAM Fitting**

**Description**

The RealVAMS package fits a multivariate value-added model (VAM) (see Broatch and Lohr 2012) with normally distributed test scores and a binary outcome indicator. This material is based upon work supported by the National Science Foundation under grants DRL-1336027 and DRL-1336265.

The package fits continuous test score results jointly with a binary outcome in a multivariate generalized linear mixed model (see Broatch and Lohr (2012); Karl, Yang, and Lohr (2013); and Karl, Yang, and Lohr (2014)) using a pseudo-likelihood approximation.

**Usage**

```r
realvams(score.data, outcome.data, persistence = "CP", school.effects = FALSE, REML = TRUE, score.fixed.effects = formula(~as.factor(year) + 0), outcome.fixed.effects = formula(~1), max.iter.EM = 10, outcome.family = binomial(link = "probit"), tol1 = 1e-07, max.PQL.it = 30, pconv = .Machine$double.eps*1e9, var.parm.hessian = TRUE, verbose = TRUE)
```

**Arguments**

- **score.data**: a data frame that contains at least a column "y" containing the student scores, a column "student" containing unique student ID's, a column "teacher" containing the teacher ID's, and a column "year" which contains the year (or semester, etc.) of the time period. The "y" and "year" variables needs to be numeric. If other variables are to be included as fixed effects, they should also be included in score.data. See 'Note' for further discussion.

- **outcome.data**: a data frame that contains at least a column "r" containing the binary student outcomes (coded 0/1), and a column "student" containing unique student ID's. The student ID's should match those in score.data. If other variables are to be included as fixed effects, they should also be included in outcome.data.

- **persistence**: a character. Choices are "CP" or "VP", for complete and variable persistence of the teacher score effects, respectively. The teacher outcome effects are modeled with complete persistence, regardless of the selection here.

- **school.effects**: logical. If TRUE, correlated random school-level effects are fitted in the score and outcome response models. For both responses, the school effects are fit with zero-persitence (a student's score in each year is associated with the current school attended, and their outcome is associated with the last school the student attended). The school ID should be included as a column schoolID in the score.data data frame.
REML logical. If TRUE, the pseudo-response is fit using REML. If FALSE, ML is used.

score.fixed.effects an object of class formula describing the structure of the fixed effects for the student scores. Categorical variables should be wrapped in an as.factor statement.

outcome.fixed.effects an object of class formula describing the structure of the fixed effects for the student outcomes. Categorical variables should be wrapped in an as.factor statement.

max.iter.EM numeric. The maximum number of EM iterations during each pseudo-likelihood iteration

outcome.family an object of class family describing the assumed distribution of the response. Currently only "binomial" has been tested, though "poisson" should work as well.

tol1 numeric. Convergence tolerance for EM algorithm during each interior pseudo-likelihood iteration. The convergence criterion is specified under 'Details'.

max.PQL.it numeric. Maximum number of outer pseudo-likelihood iterations.

pconv numeric. Convergence criterion for outer pseudo-likelihood iterations. Compare to the PCONV option of SAS PROC GLIMMIX.

var.parm.hessian logical. If TRUE, the Hessian of the parameters in the error and random effects covariance matrices is calculated, providing standard errors for those parameters. Setting this option to FALSE will reduce the run time of the program: only standard errors for the fixed effects will be returned.

verbose logical. If TRUE, model information will be printed at each iteration.

Details

*The persistence option determines the type of persistence effects that are modeled. The variable persistence model ("VP") assumes that teacher effects in future years are multiples of their effect in the current year (Lockwood et al. 2007). The multipliers in the VP model are called persistence parameters, and are estimated. By contrast, the complete persistence ("CP") model fixes the persistence parameters at 1 and 0 (Lockwood et al. 2007).

*Convergence is declared for each interior iteration when \((l_k - l_{k-1})/l_k < tol1\), where \(l_k\) is the log-likelihood at iteration \(k\).

*The model is linearized using a pseudo-likelihood approach (Wolfgang 1993) and the resulting multiple membership linear mixed model is estimated via an EM algorithm (Karl et al. 2012).

Value

RealVAMS returns an object of class RealVAMS

loglik the maximized log-likelihood at convergence of the EM algorithm. Warning: Likelihood-ratio tests are not valid with results from a PQL estimation routine.

teach.effects a data frame containing the predicted teacher effects and standard errors

parameters a matrix of estimated model parameters and standard errors
RealVAMS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hessian</td>
<td>the Hessian of the variance parameters</td>
</tr>
<tr>
<td>R_i</td>
<td>a matrix containing the error covariance matrix of a student. The bottom-right component corresponds to the variance of the binary response, and is fixed at 1.</td>
</tr>
<tr>
<td>teach.cov</td>
<td>a list containing the unique blocks of the covariance matrix of teacher effects (the G matrix)</td>
</tr>
<tr>
<td>mresid</td>
<td>a vector of the raw marginal residuals</td>
</tr>
<tr>
<td>cresid</td>
<td>a vector of the raw conditional residuals</td>
</tr>
<tr>
<td>y</td>
<td>a vector of the pseudo-responses from the final PQL iteration. The test scores will be the same as those given as an input, but the 0/1 responses for the binary distribution will be different.</td>
</tr>
<tr>
<td>yhat</td>
<td>a vector of the predicted values</td>
</tr>
<tr>
<td>num.obs</td>
<td>total number of observations (test scores and binary responses)</td>
</tr>
<tr>
<td>num.student</td>
<td>total number of students included in the data</td>
</tr>
<tr>
<td>num.year</td>
<td>number of years over which test scores were modeled</td>
</tr>
<tr>
<td>num.teach</td>
<td>a vector listing the number of teachers in each year</td>
</tr>
<tr>
<td>persistence</td>
<td>a character vector indicating the persistence structure (VP or CP) used to model the teacher test-score effects</td>
</tr>
<tr>
<td>persistence_parameters</td>
<td>a matrix of the persistence parameters. The (i,j)-th component gives the persistence parameter for year-j teachers on year-i scores.</td>
</tr>
<tr>
<td>X</td>
<td>the fixed effects design matrix</td>
</tr>
<tr>
<td>Z</td>
<td>the random effects design matrix</td>
</tr>
<tr>
<td>G</td>
<td>the random effects covariance matrix</td>
</tr>
<tr>
<td>R.full</td>
<td>the error covariance matrix, which is formed as the product diag(sqrt.w)%<em>%R%</em>%diag(sqrt.w). The matrix R assumes a variance of 1 for all of the binomial responses, while R.full includes the variance from the binomial distribution (in Wolfinger (1993), diag(sqrt.w) is called R_mu).</td>
</tr>
<tr>
<td>sqrt.w</td>
<td>vector of weights for the error covariance matrix. See the description for R.full above</td>
</tr>
</tbody>
</table>

**Note**

The first few iterations of the EM algorithm will take longer than subsequent iterations. This is a result of the hybrid gradient-ascent/Newton-Raphson method used in the M-step for the R matrix in the first two iterations (Karl et al. 2012).

The model assumes that each teacher teaches only one year. If, for example, a teacher teaches in years 1 and 2, his/her first year performance is modeled independently of the second year performance. To keep these effects separate, the program appends "(year i)" to each teacher name, where i is the year in which the teacher taught.

The `fixed.effects` arguments of RealVAMS utilizes the functionality of R's `formula` class. In the statement score.fixed.effects= formula(~as.factor(year)+cont_var+0)), as.factor(year) identifies year as a categorical variable. +0 indicates that no intercept is to be fitted, and +cont_var indicates that a separate effect is to be fitted for the continuous variable "cont_var." An interaction between "year" and "cont_var" could be specified by ~as.factor(year)*cont_var+0, or equivalently, ~as.factor(year)+cont_var+as.factor(year):cont_var+0. See `formula` for more details.
Author(s)
Andrew Karl <akarl@asu.edu> Jennifer Broatch Jennifer Green

References


Examples
```r
data(example.score.data)
data(example.outcome.data)
#The next line exists to show that the function can run and that the package
#installed correctly
res.test<RealVAMS(example.score.data,example.outcome.data,max.PQL.it=1,max.iter.EM=2,
  var.parm.hessian=FALSE)
#The next line (not run automatically) provides a full example of the function
## Not run: res<RealVAMS(example.score.data,example.outcome.data)
```

---

**R_mstep2**

**Internal function**

Description

An internal function

Usage

```r
R_mstep2(invsqrtW_,JYp_,loopsizes_,patternlength_,rownumber_,ybetas_,etahat_,tempmatR_,
  JXpi_,JXpp_,JXpx_,JXpdim_,JZpi_,JZpp_,JZpx_,JZpdim_)```

Arguments

- invsqrtW_ an internal variable
- Jyp_ an internal variable
- loopsize_ an internal variable
- patternlength_ an internal variable
- rownumber_ an internal variable
- ybetas_ an internal variable
- etahat_ an internal variable
- tempmatR_ an internal variable
- JXpi_ an internal variable
- JXpp_ an internal variable
- JXpx_ an internal variable
- JXpdim_ an internal variable
- JZpi_ an internal variable
- JZpp_ an internal variable
- JZpx_ an internal variable
- JZpdim_ an internal variable

vp_cp

Internal function

Description

An internal function

Usage

vp_cp(Z_mat, B.mat, control)

Arguments

- Z_mat data frame
- B.mat data frame
- control a list
Index

*Topic **datasets**
  example.outcome.data, 3
  example.score.data, 4

*Topic **package**
  RealVAMS-package, 2

*Topic **regression**
  RealVAMS, 5

example.outcome.data, 3
example.score.data, 4

formula, 7

R_mstep2, 8
RealVAMS, 5
RealVAMS-package, 2

vp_cp, 9