Package ‘wfe’

February 20, 2015

Version 1.3
Date 2014-08-09
Title Weighted Linear Fixed Effects Regression Models for Causal Inference
Author In Song Kim <insong@mit.edu>, Kosuke Imai <kimai@Princeton.edu>
Maintainer In Song Kim <insong@mit.edu>
Depends R (>= 2.11.0), utils, arm (>= 1.4-6), Matrix, MASS
Description This R package provides a computationally efficient way of fitting weighted linear fixed effects estimators for causal inference with various weighting schemes. Imai and Kim (2012) show that weighted linear fixed effects estimators can be used to estimate the average treatment effects under different identification strategies. This includes stratified randomized experiments, matching and stratification for observational studies, first differencing, and difference-in-differences. The package also provides various robust standard errors and a specification test for standard linear fixed effects estimators.
LazyLoad yes
LazyData yes
License GPL (>= 2)
NeedsCompilation yes
Repository CRAN
Date/Publication 2014-08-11 05:56:52

R topics documented:

  pwfe ........................................................................... 2
  wfe ........................................................................... 7

Index 12
pwfe

\textit{Fitting the Weighted Fixed Effects Model with Propensity Score Weighting}

\section*{Description}

\texttt{pwfe} is used to fit weighted fixed effects model for causal inference after transforming outcome variable based on estimated propensity score. \texttt{pwfe} also derives the regression weights for different causal quantity of interest.

\section*{Usage}

\texttt{pwfe(formula, treat = "treat.name", outcome, data, pscore = NULL, unit.index, time.index = NULL, method = "unit", within.unit = TRUE, qoi = c("ate", "att"), estimator = NULL, C.it = NULL, White = TRUE, White.alpha = 0.05, hetero.se = TRUE, auto.se = TRUE, unbiased.se = FALSE, verbose = TRUE)}

\section*{Arguments}

\begin{itemize}
  \item \texttt{formula} \hspace{1cm} a symbolic description of the model for estimating propensity score. The formula should not include dummies for fixed effects. The details of model specifications are given under ‘Details’.
  \item \texttt{treat} \hspace{1cm} a character string indicating the name of treatment variable used in the models. The treatment should be binary indicator (integer with 0 for the control group and 1 for the treatment group).
  \item \texttt{outcome} \hspace{1cm} a character string indicating the name of outcome variable.
  \item \texttt{data} \hspace{1cm} data frame containing the variables in the model.
  \item \texttt{pscore} \hspace{1cm} an optional character string indicating the name of estimated propensity score. Note that pre-specified propensity score should be bounded away from zero and one.
  \item \texttt{unit.index} \hspace{1cm} a character string indicating the name of unit variable used in the models. The index of unit should be factor.
  \item \texttt{time.index} \hspace{1cm} a character string indicating the name of time variable used in the models. The index of time should be factor.
  \item \texttt{method} \hspace{1cm} method for weighted fixed effects regression, either \texttt{unit} for unit fixed effects; \texttt{time} for time fixed effects. The default is \texttt{unit}.
  \item \texttt{within.unit} \hspace{1cm} a logical value indicating whether propensity score is estimated within unit. The default is \texttt{TRUE}.
  \item \texttt{qoi} \hspace{1cm} one of "ate" or "att". The default is "ate".
  \item \texttt{estimator} \hspace{1cm} an optional character string "fd" indicating whether the first-difference estimator will be used.
\end{itemize}
C.it an optional non-negative numeric vector specifying relative weights for each unit of analysis.

White a logical value indicating whether White misspecification statistics should be calculated. The default is TRUE.

White.alpha level of functional specification test. See White (1980) and Imai. The default is 0.05.

hetero.se a logical value indicating whether heteroskedasticity across units is allowed in calculating standard errors. The default is TRUE.

auto.se a logical value indicating whether arbitrary autocorrelation is allowed in calculating standard errors. The default is TRUE.

unbiased.se logical. If TRUE, bias-adjusted heteroskedasticity-robust standard errors are used. See Stock and Watson (2008). Should be used only for balanced panel. The default is FALSE.

verbose logical. If TRUE, helpful messages along with a progress report of the weight calculation are printed on the screen. The default is TRUE.

Details

To fit the weighted unit (time) fixed effects model with propensity score weighting, use the syntax for the formula, ~ x1 + x2, where x1 and x2 are unit (time) varying covariates.

One can provide his/her own estimated pscore which can be used to transform the outcome variable. If so, one does not need to specify formula.

If pscore is not provided, bayesglm will be used to estimate propensity scores. If within.unit = TRUE, propensity score will be separately estimated within time (unit) when method is unit (time). Otherwise, propensity score will be estimated on entire data at once.

The estimated propensity scores will be used to transform the outcome variable as described in Imai and Kim (2011).

pwfe calculates weights based on different underlying causal quantity of interest: Average Treatment Effect (qoi = "ate") or Average Treatment Effect for the Treated (qoi = "att").

One can further set estimating methods: First-Difference (estimator = "fd") or Difference-in-differences (estimator = "did").

To specify different ex-ante weights for each unit of analysis, use non-negative weights C.it. For instance, using the survey weights for C.it enables the estimation of the average treatment effect for the target population.

Value

pwfe returns an object of class "pwfe", a list that contains the components listed below.

The function summary (i.e., summary.pwfe) can be used to obtain a table of the results.

coefficients a named vector of coefficients
residuals the residuals, that is respons minus fitted values
df the degree of freedom
W weight matrix calculated from the model. Row and column indices can be found from unit.name, time.name.
call the matched call
causal causal quantity of interest
estimator the estimating method
unit.name a vector containing unique unit names
unit.index a vector containing unique unit index number
time.name a vector containing unique time names
time.index a vector containing unique time index number
method call of the method used
vcov the variance covariance matrix
White.alpha the alpha level for White specification test
White.pvalue the p-value for White specification test
White.stat the White statistics
x the design matrix
y the response vector
mf the model frame

Author(s)
Kosuke Imai, Princeton University, <kimai@princeton.edu> and In Song Kim, Princeton University, <insong@princeton.edu>

References


See Also
wfe for fitting weighted fixed effect models.

Examples
### NOTE: this example illustrates the use of wfe function with randomly generated panel data with arbitrary number of units and time.

## generate panel data with number of units = N, number of time = Time
N <- 10 # number of distinct units
Time <- 15 # number of distinct time

## generate treatment variable
```r
treat <- matrix(rbinom(N*Time, size = 1, 0.25), ncol = N)
## make sure at least one observation is treated for each unit
while ( (sum(apply(treat, 2, mean) == 0) > 0) | (sum(apply(treat, 1, mean) == 0) > 0) | (sum(apply(treat, 2, mean) == 1) > 0) | (sum(apply(treat, 1, mean) == 1) > 0) ) {
  treat <- matrix(rbinom(N*Time, size = 1, 0.25), ncol = N)
}

treat.vec <- c(treat)

## unit fixed effects
alphai <- rnorm(N, mean = apply(treat, 2, mean))

## generate two random covariates
x1 <- matrix(rnorm(N*Time, 0.5,1), ncol=N)
x2 <- matrix(rbeta(N*Time, 5,1), ncol=N)
pscore <- matrix(runif(N*Time, 0,1), ncol=N)
x1.vec <- c(x1)
x2.vec <- c(x2)
pscore <- c(pscore)

## generate outcome variable
y <- matrix(NA, ncol = N, nrow = Time)
for (i in 1:N) {
  y[, i] <- alphai[i] + treat[, i] + x1[, i] + x2[, i] + rnorm(Time)
}
y.vec <- c(y)

## generate unit and time index
unit.index <- rep(1:N, each = Time)
time.index <- rep(1:Time, N)

Data.str <- as.data.frame(cbind(y.vec, treat.vec, unit.index, x1.vec, x2.vec))
colnames(Data.str) <- c(“y”, “tr”, “strata.id”, “x1”, “x2”)

Data.obs <- as.data.frame(cbind(y.vec, treat.vec, unit.index, time.index, x1.vec, x2.vec, pscore))
colnames(Data.obs) <- c(“y”, “tr”, “unit”, “time”, “x1”, “x2”, “pscore”)

###############################################################
# Example 1: Stratified Randomized Experiments
###############################################################

# run the weighted fixed effect regression with strata fixed effect.
# Note: the quantity of interest is Average Treatment Effect ("ate")
# and the standard errors allow heteroskedasticity and arbitrary
# autocorrelation.

### Average Treatment Effect
ps.ate <- pwfe(~ x1+x2, treat = "tr", outcome = "y", data = Data.str,
  unit.index = "strata.id", method = "unit", within.unit = TRUE,
  qoi = "ate", hetero.se=TRUE, auto.se=TRUE)

# summarize the results
summary(ps.ate)
```
### Average Treatment Effect for the Treated

```r
pwfe(x1+x2, treat = "tr", outcome = "y", data = Data.str,
   unit.index = "strata.id", method = "unit", within.unit = TRUE,
   qoi = "att", hetero.se=TRUE, auto.se=TRUE)
```

# summarize the results

```r
summary(ps.att)
```

### Example 2: Observational Studies with Unit Fixed-effects

## run the weighted fixed effect regression with unit fixed effect.

**Note:** the quantity of interest is **Average Treatment Effect ("ate")**

### Average Treatment Effect

```r
ps.obs <- pwfe(x1+x2, treat = "tr", outcome = "y", data = Data.obs,
   unit.index = "unit", time.index = "time",
   method = "unit", within.unit = TRUE,
   qoi = "ate", hetero.se=TRUE, auto.se=TRUE)
```

# summarize the results

```r
summary(ps.obs)
```

# extracting weights

```r
summary(ps.obs)$Weights
```

### Average Treatment Effect with First-difference

```r
ps.fd <- pwfe(x1+x2, treat = "tr", outcome = "y", data = Data.obs,
   unit.index = "unit", time.index = "time",
   method = "unit", within.unit = TRUE,
   qoi = "ate", estimator = "fd", hetero.se=TRUE, auto.se=TRUE)
```

# summarize the results

```r
summary(ps.fd)
```

### Example 3: Estimation with pre-specified propensity score

### Average Treatment Effect with Pre-specified Propensity Scores

```r
mod.ps <- pwfe(treat = "tr", outcome = "y", data = Data.obs, pscore = "pscore",
   unit.index = "unit", time.index = "time",
   method = "unit", within.unit = TRUE,
   qoi = "ate", hetero.se=TRUE, auto.se=TRUE)
```

# summarize the results
**Fitting the Weighted Fixed Effects Model for Causal Inference**

**Description**

`wfe` is used to fit weighted fixed effects model for causal inference. `wfe` also derives the regression weights for different causal quantity of interest.

**Usage**

```r
wfe(formula, data, treat = "treat.name",  
     unit.index, time.index = NULL, method = "unit",  
     qoi = "ate", estimator = NULL, C.it = NULL,  
     hetero.se = TRUE, auto.se = TRUE,  
     White = TRUE, White.alpha = 0.05,  
     verbose = TRUE, unbiased.se = FALSE, unweighted = FALSE,  
     store.wdm = FALSE,  
     tol = sqrt(.Machine$double.eps))
```

**Arguments**

- `formula` a symbolic description of the model to be fitted. The formula should not include dummies for fixed effects. The details of model specifications are given under 'Details'.
- `data` data frame containing the variables in the model.
- `treat` a character string indicating the name of treatment variable used in the models. The treatment should be binary indicator (integer with 0 for the control group and 1 for the treatment group).
- `unit.index` a character string indicating the name of unit variable used in the models. The index of unit should be factor.
- `time.index` a character string indicating the name of time variable used in the models. The index of time should be factor.
- `method` method for weighted fixed effects regression, either `unit` for unit fixed effects; `time` for time fixed effects. The default is `unit`.
- `qoi` one of "ate" or "att". The default is "ate".
- `estimator` an optional character string indicating the estimating method. One of "fd" or "did". The default is `NULL`.
- `C.it` an optional non-negative numeric vector specifying relative weights for each unit of analysis.
- `hetero.se` a logical value indicating whether heteroskedasticity across units is allowed in calculating standard errors. The default is `TRUE`. 
auto.se a logical value indicating whether arbitrary autocorrelation is allowed in calculating standard errors. The default is TRUE.

White a logical value indicating whether White misspecification statistics should be calculated. The default is TRUE.

White.alpha level of functional specification test. See White (1980) and Imai and Kim (2012). The default is 0.05.

verbose logical. If TRUE, helpful messages along with a progress report of the weight calculation are printed on the screen. The default is TRUE.

unbiased.se logical. If TRUE, bias-adjusted heteroskedasticity-robust standard errors are used. See Stock and Watson (2008). Should be used only for balanced panel. The default is FALSE.

unweighted logical. If TRUE, standard unweighted fixed effects model is estimated. The default is FALSE. Note: users do not need to specify qoi when unweighted=TRUE.

store.wdm logical. If TRUE, weighted demeaned dataframe will be stored. The default is FALSE.

tol a relative tolerance to detect zero singular values for generalized inverse. The default is $\sqrt{\text{.Machine\$double.eps}}$.

Details
To fit the weighted unit (time) fixed effects model, use the syntax for the formula, \( y \sim x_1 + x_2 \), where \( y \) is a dependent variable and \( x_1 \) and \( x_2 \) are unit (time) varying covariates.

\texttt{wfe}} calculates weights based on different underlying causal quantity of interest: Average Treatment Effect (qoi = "ate") or Average Treatment Effect for the Treated (qoi = "att").

One can further set estimating methods: First-Difference (estimator = "fd") or Difference-in-differences (estimator = "did"). For the two-way fixed effects model, set estimator = "did"

To specify different ex-ante weights for each unit of analysis, use non-negative weights \texttt{C.it}. For instance, using the survey weights for \texttt{C.it} enables the estimation for the average treatment effect for the target population.

An object of class "wfe" contains vectors of unique unit(time) names and unique unit(time) indices.

Value
\texttt{wfe} returns an object of class "wfe", a list that contains the components listed below.

The function \texttt{summary} (i.e., \texttt{summary.wfe}) can be used to obtain a table of the results.

\begin{itemize}
  \item \texttt{coefficients} a named vector of coefficients
  \item \texttt{residuals} the residuals, that is respons minus fitted values
  \item \texttt{df} the degree of freedom
  \item \texttt{W} weight matrix calculated from the model. Row and column indices can be found from \texttt{units.times}
  \item \texttt{call} the matched call
  \item \texttt{causal} causal quantity of interest
\end{itemize}
estimator  the estimating method
units      a dataframe containing unit names used for \( W \)
times      a dataframe containing time names used for \( W \)
method     call of the method used
vcov       the variance covariance matrix
White.alpha the alpha level for White specification test
White.pvalue the p-value for White specification test
White.stat  the White statistics
X           the design matrix
Y           the response vector
X.wdm       the demeaned design matrix
Y.wdm       the demeaned response vector
mf          the model frame

Author(s)
Kosuke Imai, Princeton University, <kimai@princeton.edu> and In Song Kim, Princeton University, <insong@princeton.edu>

References

See Also
pwfe for fitting weighted fixed effects models with propensity score weighting

Examples
### NOTE: this example illustrates the use of wfe function with randomly
generated panel data with arbitrary number of units and time.

## generate panel data with number of units = N, number of time = Time
N <- 10 # number of distinct units
Time <- 15 # number of distinct time

## treatment effect
beta <- 1

## generate treatment variable
```r
treat <- matrix(rbinom(N*Time, size = 1, 0.25), ncol = N)
## make sure at least one observation is treated for each unit
while ((sum(apply(treat, 2, mean) == 0) > 0) | (sum(apply(treat, 2, mean) == 1) > 0) |
  (sum(apply(treat, 1, mean) == 0) > 0) | (sum(apply(treat, 1, mean) == 1) > 0)) {
  treat <- matrix(rbinom(N*Time, size = 1, 0.25), ncol = N)
}
treat.vec <- c(treat)

## unit fixed effects
alphai <- rnorm(N, mean = apply(treat, 2, mean))

## generate two random covariates
x1 <- matrix(rnorm(N*Time, 0.5,1), ncol=N)
x2 <- matrix(rbeta(N*Time, 5,1), ncol=N)
x1.vec <- c(x1)
x2.vec <- c(x2)

## generate outcome variable
y <- matrix(NA, ncol = N, nrow = Time)
for (i in 1:N) {
  y[i, i] <- alphai[i] + treat[i] + x1[,i] + x2[,i] + rnorm(Time)
}
y.vec <- c(y)

## generate unit and time index
unit.index <- rep(1:N, each = Time)
time.index <- rep(1:Time, N)

Data.str <- as.data.frame(cbind(y.vec, treat.vec, unit.index, x1.vec, x2.vec))
colnames(Data.str) <- c("y", "tr", "strata.id", "x1", "x2")

Data.obs <- as.data.frame(cbind(y.vec, treat.vec, unit.index, time.index, x1.vec, x2.vec))
colnames(Data.obs) <- c("y", "tr", "unit", "time", "x1", "x2")

# Example 1: Stratified Randomized Experiments
# Note: the quantity of interest is Average Treatment Effect ("ate")
# and the standard errors allow heteroskedasticity and arbitrary
# autocorrelation.

### Average Treatment Effect
mod.ate <- wfe(y~ tr+x1+x2, data = Data.str, treat = "tr",
  unit.index = "strata.id", method = "unit",
  qoi = "ate", hetero.se=TRUE, auto.se=TRUE)

### summarize the results
summary(mod.ate)

### Average Treatment Effect for the Treated
mod.att <- wfe(y~ tr+x1+x2, data = Data.str, treat = "tr",
  unit.index = "strata.id", method = "unit",
  qoi = "ate", hetero.se=TRUE, auto.se=TRUE)
```
wfe

qoi = "att", hetero.se=TRUE, auto.se=TRUE)

## summarize the results
summary(mod.att)

# Example 2: Observational Studies with Unit Fixed-effects
# run the weighted fixed effect regression with unit fixed effect.
# Note: the quantity of interest is Average Treatment Effect ("ate")
# and the standard errors allow heteroskedasticity and arbitrary 
# autocorrelation.

mod.obs <- wfe(y ~ tr+x1+x2, data = Data.obs, treat = "tr",
               unit.index = "unit", time.index = "time", method = "unit",
               qoi = "ate", hetero.se=TRUE, auto.se=TRUE,
               White = TRUE, White.alpha = 0.05)

## summarize the results
summary(mod.obs)

## extracting weigths
summary(mod.obs)$W

# Example 3: Observational Studies with differences-in-differences
# run difference-in-differences estimator.
# Note: the quantity of interest is Average Treatment Effect ("ate")
# and the standard errors allow heteroskedasticity and arbitrary 
# autocorrelation.

mod.did <- wfe(y ~ tr+x1+x2, data = Data.obs, treat = "tr",
               unit.index = "unit", time.index = "time", method = "unit",
               qoi = "ate", estimator ="did", hetero.se=TRUE, auto.se=TRUE,
               White = TRUE, White.alpha = 0.05, verbose = TRUE)

## summarize the results
summary(mod.did)

## extracting weigths
summary(mod.did)$W
Index

*Topic regression
  pwfe, 2
  wfe, 7

pwfe, 2
wfe, 7