

# Package: agfh (via r-universe)

September 17, 2024

**Type** Package

**Version** 0.2.1

**Date** 2023-06-21

**Title** Agnostic Fay-Herriot Model for Small Area Statistics

**Description** Implements the Agnostic Fay-Herriot model, an extension of the traditional small area model. In place of normal sampling errors, the sampling error distribution is estimated with a Gaussian process to accommodate a broader class of distributions. This flexibility is most useful in the presence of bounded, multi-modal, or heavily skewed sampling errors.

**License** GPL ( $\geq 3$ )

**Encoding** UTF-8

**Imports** ggplot2, goftest, ks, mvtnorm, stats

**Suggests** knitr, rmarkdown, testthat ( $\geq 3.0.0$ )

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Marten Thompson [aut, cre, cph], Snigdhansu Chatterjee [ctb, cph]

**Maintainer** Marten Thompson <thom7058@umn.edu>

**Date/Publication** 2023-06-21 20:00:05 UTC

**Repository** <https://martenthompson.r-universe.dev>

**RemoteUrl** <https://github.com/cran/agfh>

**RemoteRef** HEAD

**RemoteSha** 61b98cccbc808a45e565c8888618cc8d84859a67

## Contents

adj_profile_likelihood_theta_var_maker . . . . .	2
adj_resid_likelihood_theta_var_maker . . . . .	3

agfh_theta_new_pred . . . . .	4
anderson_darling . . . . .	5
beta_err_gen . . . . .	6
cramer_vonmises . . . . .	7
gamma_err_gen . . . . .	8
hb_theta_new_pred . . . . .	9
kolmogorov_smirnov . . . . .	10
make_agfh_sampler . . . . .	11
make_gibbs_sampler . . . . .	12
map_from_density . . . . .	13
mse . . . . .	13
null_gen . . . . .	14
resid_likelihood_theta_var_maker . . . . .	15
RM_beta_eblue . . . . .	16
RM_theta_eblup . . . . .	17
RM_theta_new_pred . . . . .	18
RM_theta_var_moment_est . . . . .	18
shapiro_wilk . . . . .	19
test_u_normal . . . . .	20
theta_var_est_grid . . . . .	21

<b>Index</b>	<b>22</b>
--------------	-----------

---

adj\_profile\_likelihood\_theta\_var\_maker

*Maker Function: Adjusted Profile Likelihood of Latent Variance*

---

## Description

A maker function that returns a function. The returned function is the adjusted profile likelihood of the data for a given (latent) variance, from Yoshimori & Lahiri (2014).

## Usage

```
adj_profile_likelihood_theta_var_maker(X, Y, D)
```

## Arguments

X	observed independent data to be analyzed
Y	observed dependent data to be analyzed
D	known precisions of response Y

## Value

Returns the adjusted profile likelihood as a function of the variance term in the latent model.

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
X <- matrix(1:10, ncol=1)
Y <- 2*X + rnorm(10, sd=1.1)
D <- rep(1, 10)
adj.lik <- adj_profile_likelihood_theta_var_maker(X, Y, D)
adj.lik(0.5)
```

---

adj\_resid\_likelihood\_theta\_var\_maker

*Maker Function: Adjusted Residual Likelihood of Latent Variance*

---

**Description**

A maker function that returns a function. The returned function is the adjusted residual likelihood of the data for a given (latent) variance, from Yoshimori & Lahiri (2014).

**Usage**

```
adj_resid_likelihood_theta_var_maker(X, Y, D)
```

**Arguments**

X	observed independent data to be analyzed
Y	observed dependent data to be analyzed
D	known precisions of response Y

**Value**

Returns the adjusted residual likelihood as a function of the variance term in the latent model.

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
X <- matrix(1:10, ncol=1)
Y <- 2*X + rnorm(10, sd=1.1)
D <- rep(1, 10)
adj.lik <- adj_resid_likelihood_theta_var_maker(X, Y, D)
adj.lik(0.5)
```

---

agfh\_theta\_new\_pred    *Agnostic Fay-Herriot Hierarchical Bayesian Model Predictions at Latent Level*

---

### Description

Find predictions of  $\theta$  using posterior samples from the AGFH model

### Usage

```
agfh_theta_new_pred(X_new, beta_samples, theta_var_samples)
```

### Arguments

`X_new`                    single new independent data to be analyzed  
`beta_samples`            posterior samples of latent regression parameter  
`theta_var_samples`        posterior samples of latent variance parameter

### Details

`X_new` should be  $1 \times p$  shaped.

`beta_samples` and `theta_var_samples` should contain the same number of samples (columns for the former, length of the latter).

### Value

Vector containing  $n$  samples-many estimates of  $\theta$  at `X_new`.

### Source

Marten Thompson [thom7058@umn.edu](mailto:thom7058@umn.edu)

### Examples

```
p <- 3
n.post.samp <- 10
X.new <- matrix(rep(1,p), nrow=1)
beta.samp <- matrix(rnorm(n.post.samp*p, mean=2, sd=0.1), ncol=n.post.samp)
thvar.samp <- runif(n.post.samp, 0.1, 1)

th.preds <- agfh_theta_new_pred(X.new, beta.samp, thvar.samp)
```

---

anderson_darling	<i>Anderson-Darling Normality Test</i>
------------------	--

---

**Description**

Test a sample against the null hypothesis that it comes from a standard Normal distribution.

**Usage**

```
anderson_darling(samples)
```

**Arguments**

samples            vector of values to be tested

**Details**

Wrapper function for corresponding functionality in `gof.test`. Originally, from Anderson and Darling (1954).

**Value**

A list containing

name                authors of normality test applied i.e. 'Anderson Darling'

statistic           scalar value of test statistics

p.value             corresponding p-value of the test

**Source**

Anderson and Darling (1954) via `gof.test`.

**Examples**

```
sample <- rnorm(100)
anderson_darling(sample)
```

---

beta_err_gen	<i>Generate Data with Beta Sampling Errors</i>
--------------	--

---

### Description

The traditional Fay-Herriot small area model has a Normal latent variable and Normal observed response errors. This method generates data with Normal latent variables and Beta errors on the response. Note that the sampling errors are transformed so their mean and variance match the the first two moments of the traditional model.

### Usage

```
beta_err_gen (M, p, D, lambda, a, b)
```

### Arguments

M	number of areal units
p	dimension of regressors i.e. $x \in R^p$
D	vector of precisions for response, length M
lambda	value of latent variance
a	first shape parameter of Beta distribution
b	second shape parameter of Beta distribution

### Value

A list containing

D	copy of argument ‘D’
beta	vector of length ‘p’ latent coefficients
lambda	copy of argument ‘lambda’
X	matrix of independent variables
theta	vector of latent effects
Y	vector of responses
err	vector of sampling errors
name	name of sampling error distribution, including shape parameters

### Source

Marten Thompson [thom7058@umn.edu](mailto:thom7058@umn.edu)

### Examples

```
M <- 50
p <- 3
D <- rep(0.1, M)
lamb <- 1/2
dat <- beta_err_gen(M, p, D, lamb, 1/2, 1/4)
```

---

cramer_vonmises	<i>Cramer-Von Mises Normality Test</i>
-----------------	--

---

**Description**

Test a sample against the null hypothesis that it comes from a standard Normal distribution.

**Usage**

```
cramer_vonmises(samples)
```

**Arguments**

`samples` vector of values to be tested

**Details**

Wrapper function for corresponding functionality in `gof test`. Originally developed in Cramer (1928), Mises (1931), and Smirnov (1936).

**Value**

A list containing

`name` authors of normality test applied i.e. 'Cramer von Mises'

`statistic` scalar value of test statistics

`p.value` corresponding p-value of the test

**Source**

Cramer (1928), Mises (1931), and Smirnov (1936) via `gof test`.

**Examples**

```
sample <- rnorm(100)
cramer_vonmises(sample)
```

---

 gamma\_err\_gen

*Generate Data with Gamma Sampling Errors*


---

### Description

The traditional Fay-Herriot small area model has a Normal latent variable and Normal observed response errors. This method generates data with Normal latent variables and Gamma errors on the response. Note that the sampling errors are transformed so their mean and variance match the the first two moments of the traditional model.

### Usage

```
gamma_err_gen (M, p, D, lambda, shape, rate)
```

### Arguments

M	number of areal units
p	dimension of regressors i.e. $x \in R^p$
D	vector of precisions for response, length M
lambda	value of latent variance
shape	shape parameter of Gamma distribution
rate	rate parameter of Gamma distribution

### Value

A list containing

D	copy of argument ‘D’
beta	vector of length ‘p’ latent coefficients
lambda	copy of argument ‘lambda’
X	matrix of independent variables
theta	vector of latent effects
Y	vector of responses
err	vector of sampling errors
name	name of sampling error distribution, including shape and rate parameters

### Source

Marten Thompson thom7058@umn.edu

### Examples

```
M <- 50
p <- 3
D <- rep(0.1, M)
lamb <- 1/2
dat <- gamma_err_gen(M, p, D, lamb, 1/2, 10)
```



---

hb\_theta\_new\_pred      *Traditional Fay-Herriot Hierarchical Bayesian Model Predictions*

---

### Description

Find predictions using posterior samples from the traditional Fay-Herriot hierarchical bayesian model

### Usage

```
hb_theta_new_pred(X_new, beta_samples, theta_var_samples)
```

### Arguments

X\_new                    single new independent data to be analyzed  
beta\_samples            posterior samples of latent regression parameter  
theta\_var\_samples       posterior samples of latent variance parameter

### Details

X\_new should be  $1 \times p$  shaped.

beta\_samples and theta\_var\_samples should contain the same number of samples (columns for the former, length of the latter).

### Value

Vector containing n samples-many estimates of  $\theta$  at X\_new.

### Source

Marten Thompson thom7058@umn.edu

### Examples

```
p <- 3
n.post.samp <- 10
X.new <- matrix(rep(1,p), nrow=1)
beta.samp <- matrix(rnorm(n.post.samp*p, mean=2, sd=0.1), ncol=n.post.samp)
thvar.samp <- runif(n.post.samp, 0.1, 1)

th.preds <- hb_theta_new_pred(X.new, beta.samp, thvar.samp)
```

---

kolmogorov\_smirnov      *Kolmogorov-Smirnov Normality Test*

---

### Description

Test a sample against the null hypothesis that it comes from a standard Normal distribution.

### Usage

```
kolmogorov_smirnov(samples)
```

### Arguments

samples            vector of values to be tested

### Details

Wrapper function for corresponding functionality in stats. Originally, from Kolmogorov (1933).

### Value

A list containing

name                name of normality test applied i.e. 'Komogorov Smirnov'

statistic           scalar value of test statistics

p.value             corresponding p-value from test

### Source

Kolmogorov (1933) via stats.

### Examples

```
sample <- rnorm(100)
kolmogorov_smirnov(sample)
```

---

make\_agfh\_sampler      *Maker Function: Agnostic Fay-Herriot Sampler*

---

### Description

A maker function that returns a function. The returned function is a sampler for the agnostic Fay-Herriot model.

### Arguments

X	observed independent data to be analyzed
Y	observed dependent data to be analyzed
D	known precisions of response Y
var_gamma_a	latent variance prior parameter, rgamma shape
var_gamma_b	latent variance prior parameter, rgamma rate
S	vector of starting support values for $g(\cdot)$
kern.a0	scalar variance parameter of GP kernel
kern.a1	scalar lengthscale parameter of GP kernel
kern.fuzz	scalar noise variance of kernel

### Details

Creates a Metropolis-within-Gibbs sampler of the agnostic Fay-Herriot model (AGFH).

### Value

Returns a sampler, itself a function of initial parameter values (a list with values for  $\beta$ ,  $\theta$ , the latent variance of  $\theta$ , and starting values for  $g(\cdot)$ , typically zeros), number of samples, thinning rate, and scale of Metropolis-Hastings jumps for  $\theta$  sampling.

### Source

Marten Thompson [thom7058@umn.edu](mailto:thom7058@umn.edu)

### Examples

```
n <- 10
X <- matrix(1:n, ncol=1)
Y <- 2*X + rnorm(n, sd=1.1)
D <- rep(1, n)
ag <- make_agfh_sampler(X, Y, D)

params.init <- list(
  beta=1,
  theta=rep(0,n),
  theta.var=1,
```

```

    gamma=rep(0,n)
  )
  ag.out <- ag(params.init, 5, 1, 0.1)

```

---

make\_gibbs\_sampler      *Maker Function: Traditional Fay-Herriot Gibbs Sampler*

---

### Description

A maker function that returns a function. The returned function is a Gibbs sampler for the traditional Fay-Herriot model.

### Usage

```
make_gibbs_sampler(X, Y, D, var_gamma_a=1, var_gamma_b=1)
```

### Arguments

X	observed independent data to be analyzed
Y	observed dependent data to be analyzed
D	known precisions of response Y
var_gamma_a	latent variance prior parameter, rgamma shape
var_gamma_b	latent variance prior parameter, rgamma rate

### Value

Returns a Gibbs sampler, itself a function of initial parameter values (a list with values for  $\beta$ ,  $\theta$ , and latent variance of  $\theta$ ), number of samples, and thinning rate.

### Source

Marten Thompson [thom7058@umn.edu](mailto:thom7058@umn.edu)

### Examples

```

n <- 10
X <- matrix(1:n, ncol=1)
Y <- 2*X + rnorm(n, sd=1.1)
D <- rep(1, n)
gib <- make_gibbs_sampler(X, Y, D)

params.init <- list(
  beta=1,
  theta=rep(0,n),
  theta.var=1
)
gib.out <- gib(params.init, 5, 1)

```

---

map_from_density	<i>Calculate the MAP Estimate from Posterior Samples</i>
------------------	--

---

**Description**

Find maximum a posteriori estimate using posterior samples

**Usage**

```
map_from_density(param.ts, plot=FALSE)
```

**Arguments**

param.ts	vector of scalar samples
plot	boolean, plot or not

**Details**

Finds location of max of density from samples.

**Value**

Scalar MAP estimate.

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
n.post.samp <- 10  
beta.samp <- rnorm(n.post.samp, 0, 1/2)  
  
map_from_density(beta.samp)
```

---

mse	<i>Calculate the Mean Squared Error Between two Vectors</i>
-----	---

---

**Description**

Merely wanted to use such a function by name; nothing fancy

**Usage**

```
mse(x,y)
```

**Arguments**

x	vector of values
y	vector of values

**Value**

A scalar: the MSE between x and y.

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
mse(seq(1:10), seq(10:1))
```

---

null_gen	<i>Generate Data with Normal Sampling Errors</i>
----------	--

---

**Description**

The Fay-Herriot small area model has a Normal latent variable and Normal observed response. This generates data according to that specification.

**Usage**

```
null_gen (M, p, D, lambda)
```

**Arguments**

M	number of areal units
p	dimension of regressors i.e. $x \in R^p$
D	vector of precisions for response, length M
lambda	value of latent variance

**Value**

A list containing

D	copy of argument ‘D’
beta	vector of length ‘p’ latent coefficients
lambda	copy of argument ‘lambda’
X	matrix of independent variables
theta	vector of latent effects
Y	vector of responses
err	vector of sampling errors
name	name of sampling error distribution

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
M <- 50
p <- 3
D <- rep(0.1, M)
lamb <- 1/2
dat <- null_gen(M, p, D, lamb)
```

---

resid\_likelihood\_theta\_var\_maker

*Maker Function: Residual Likelihood of Latent Variance*

---

**Description**

A maker function that returns a function. The returned function is the (non-adjusted) residual likelihood of the data for a given (latent) variance, from Yoshimori & Lahiri (2014).

**Usage**

```
resid_likelihood_theta_var_maker(X, Y, D)
```

**Arguments**

X	observed independent data to be analyzed
Y	observed dependent data to be analyzed
D	known precisions of response Y

**Value**

Returns the (non-adjusted) residual likelihood as a function of the variance term in the latent model.

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
X <- matrix(1:10, ncol=1)
Y <- 2*X + rnorm(10, sd=1.1)
D <- rep(1, 10)
resid.lik <- resid_likelihood_theta_var_maker(X, Y, D)
resid.lik(0.5)
```

RM\_beta\_eblue

*Traditional EBLUE Estimator of Beta*

---

**Description**

Traditional EBLUE Estimator of Beta

**Usage**

```
RM_beta_eblue(X, Y, D, theta_var_est)
```

**Arguments**

X	observed independent data to be analyzed
Y	observed dependent data to be analyzed
D	known precisions of response Y
theta_var_est	estimate of variance term for latent model

**Details**

Traditional EBLUE estimator of beta.

**Value**

Returns a vector estimate of beta.

**Source**

Marten Thompson [thom7058@umn.edu](mailto:thom7058@umn.edu)

**Examples**

```
X <- matrix(1:10, ncol=1)
Y <- 2*X + rnorm(10, sd=1.1)
D <- rep(1, 10)
th.var.est <- 0.1
RM_beta_eblue(X, Y, D, th.var.est)
```



---

RM\_theta\_eblup      *Traditional EBLUP Estimator of Theta*

---

**Description**

Traditional EBLUP Estimator of Theta

**Usage**

```
RM_theta_eblup(X, Y, D, theta.var.est)
```

**Arguments**

X	observed independent data to be analyzed
Y	observed dependent data to be analyzed
D	known precisions of response Y
theta.var.est	estimate of variance term for latent model; if NA, will automatically use method-of-moments

**Details**

Traditional EBLUP estimator of latent values theta.

**Value**

Returns a vector of estimates of theta.

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
X <- matrix(1:10, ncol=1)
Y <- 2*X + rnorm(10, sd=1.1)
D <- rep(1, 10)
th.var.est <- 0.1
RM_theta_eblup(X, Y, D, th.var.est)

RM_theta_eblup(X, Y, D)
```

---

RM\_theta\_new\_pred      *Traditional EBLUP Estimator of Theta for new X values*

---

**Description**

Traditional EBLUP Estimator of Theta for new X values

**Usage**

```
RM_theta_new_pred(X.new, beta.est)
```

**Arguments**

X.new	new independent data to be analyzed
beta.est	estimate of regression term for latent model

**Details**

Simply  $X'beta.est$

**Value**

Returns a vector of estimates of theta.

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
X <- matrix(1:10, ncol=1)
b <- 1
RM_theta_new_pred(X, b)
```

---

RM\_theta\_var\_moment\_est  
*Moment-Based Estimator of Latent Model Variance*

---

**Description**

Simple moment-based estimator of the variance of the latent model.

**Usage**

```
RM_theta_var_moment_est(X, Y, D)
```

**Arguments**

X	observed independent data to be analyzed
Y	observed dependent data to be analyzed
D	known precisions of response Y

**Details**

Simple moment-based estimator of the variance of the latent model.

**Value**

Returns a scalar estimate of variance.

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
X <- matrix(1:10, ncol=1)
Y <- 2*X + rnorm(10, sd=1.1)
D <- rep(1, 10)
RM_theta_var_moment_est(X, Y, D)
```

---

shapiro\_wilk

*Shapiro-Wilk Normality Test*

---

**Description**

Test a sample against the null hypothesis that it comes from a standard Normal distribution.

**Usage**

```
shapiro_wilk(samples)
```

**Arguments**

samples	vector of values to be tested
---------	-------------------------------

**Details**

Wrapper function for corresponding functionality in stats. Originally, from Shapiro and Wilk (1975).

**Value**

A list containing

name	authors of normality test applied i.e. 'Shapiro Wilk'
statistic	scalar value of test statistics
p.value	corresponding p-value of the test

**Source**

Shapiro and Wilk (1975) via stats.

**Examples**

```
sample <- rnorm(100)
shapiro_wilk(sample)
```

---

test_u_normal	<i>Normality Test</i>
---------------	-----------------------

---

**Description**

Test a sample against the null hypothesis that it comes from a standard Normal distribution with the specified test.

**Usage**

```
test_u_normal(samples, test)
```

**Arguments**

samples	vector of values to be tested
test	name of test, one of 'SW', 'KS', 'CM', 'AD'

**Details**

Convenience function for consistent syntax in calling shapiro\_wilk, kolmogorov\_smirnov, cramer\_vonmises, and anderson\_darling tests.

**Value**

A list containing

name	authors of normality test applied
statistic	scalar value of test statistics
p.value	corresponding p-value from test

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
sample <- rnorm(100)
test_u_normal(sample, 'SW')
```

---

theta\_var\_est\_grid      *Basic Grid Optimizer for Likelihood*

---

**Description**

A basic grid search optimizer. Here, used to estimate the variance in the latent model by maximum likelihood.

**Usage**

```
theta_var_est_grid(likelihood_theta_var)
```

**Arguments**

likelihood\_theta\_var  
some flavor of likelihood function in terms of latent variance

**Details**

likelihood\_theta\_var may be created using adj\_resid\_likelihood\_theta\_var\_maker or similar.

We recommended implementing a more robust optimizer.

**Value**

The scalar value that optimizes likelihood\_theta\_var, or an error if this value is on the search boundary  $[10^{-6}, 10^2]$ .

**Source**

Marten Thompson thom7058@umn.edu

**Examples**

```
X <- matrix(1:10, ncol=1)
Y <- 2*X + rnorm(10, sd=1.1)
D <- rep(1, 10)
adj.lik <- adj_resid_likelihood_theta_var_maker(X, Y, D)
theta_var_est_grid(adj.lik)
```

# Index

adj\_profile\_likelihood\_theta\_var\_maker, [2](#)  
adj\_resid\_likelihood\_theta\_var\_maker, [3](#)  
agfh\_theta\_new\_pred, [4](#)  
anderson\_darling, [5](#)  
  
beta\_err\_gen, [6](#)  
  
cramer\_vonmises, [7](#)  
  
gamma\_err\_gen, [8](#)  
  
hb\_theta\_new\_pred, [9](#)  
  
kolmogorov\_smirnov, [10](#)  
  
make\_agfh\_sampler, [11](#)  
make\_gibbs\_sampler, [12](#)  
map\_from\_density, [13](#)  
mse, [13](#)  
  
null\_gen, [14](#)  
  
resid\_likelihood\_theta\_var\_maker, [15](#)  
RM\_beta\_eblue, [16](#)  
rm\_beta\_eblue (RM\_beta\_eblue), [16](#)  
RM\_theta\_eblup, [17](#)  
rm\_theta\_eblup (RM\_theta\_eblup), [17](#)  
RM\_theta\_new\_pred, [18](#)  
rm\_theta\_new\_pred (RM\_theta\_new\_pred), [18](#)  
RM\_theta\_var\_moment\_est, [18](#)  
rm\_theta\_var\_moment\_est (RM\_theta\_var\_moment\_est), [18](#)  
  
shapiro\_wilk, [19](#)  
  
test\_u\_normal, [20](#)  
theta\_var\_est\_grid, [21](#)