

# Meta-regression in Stata

Roger M. Harbord  
Department of Social Medicine  
University of Bristol, UK  
roger.harbord@bristol.ac.uk

Julian P. T. Higgins  
MRC Biostatistics Unit  
Cambridge, UK  
julian.higgins@mrc-bsu.cam.ac.uk

**Abstract.** We present a revised version of the `metareg` command, which performs meta-analysis regression (meta-regression) on study-level summary data. The major revisions involve improvements to the estimation methods and the addition of an option to use a permutation test to estimate  $p$ -values, including an adjustment for multiple testing. We have also made additions to the output, added an option to produce a graph, and included support for the `predict` command. Stata 8.0 or above is required.

**Keywords:** sbe23\_1, metareg, meta-regression, meta-analysis, permutation test, multiple testing

## 1 Introduction

Meta-analysis regression, or meta-regression, is an extension to standard meta-analysis that investigates the extent to which statistical heterogeneity between results of multiple studies can be related to one or more characteristics of the studies (Thompson and Higgins 2002). Like meta-analysis, meta-regression is usually conducted on study-level summary data, because individual observations from all studies (often referred to as individual patient data in medical applications) are frequently not available.

Sharp (1998) introduced the `metareg` command to perform meta-regression on study-level summary data. In this article, we present a substantially updated and largely rewritten version of `metareg`. The planning and interpretation of meta-regression studies raises substantial statistical issues discussed at length elsewhere (Davey Smith, Egger, and Phillips 1997; Higgins et al. 2002; Thompson and Higgins 2002, 2005). In this article, we will concentrate on the rationale for and the implementation and interpretation of the following new features of `metareg`:

- An improved algorithm for the estimation of the between-study variance,  $\tau^2$ , by residual (restricted) maximum likelihood (REML)
- A modification to the calculation of standard errors,  $p$ -values, and confidence intervals for coefficients suggested by Knapp and Hartung (2003)
- Various enhancements to the output
- An option to produce a graph of the fitted model with a single covariate

- An option to calculate permutation-based  $p$ -values, including an adjustment for multiple testing based on the work of [Higgins and Thompson \(2004\)](#)
- Support for many of Stata's postestimation commands, including `predict`

We begin with a brief outline in section 2 of the statistical basis of meta-analysis and meta-regression, and we continue with a summary in section 3 of the relationship of `metareg` to other Stata commands. Section 4 introduces two example datasets that we use to illustrate the discussion of new features in section 5, which constitutes the main body of the article and has subsections corresponding to each of the new features listed above. The final two sections are reference material: Section 6 gives the Stata syntax and full list of options for `metareg` and `predict` after `metareg`, and lists the results saved by the command. Finally, section 7 gives details of the methods and formulas used.

## 2 Basis of meta-regression

In this section, we outline the statistical basis of random- and fixed-effects meta-regression and their relation to random- and fixed-effects meta-analysis. We will use mathematical formulas for brevity and precision. Less mathematically inclined readers or those who are already familiar with the principles of meta-analysis and meta-regression can skip this section.

We assume that study  $i$  of a total of  $n$  studies provides an estimate,  $y_i$ , of the effect of interest, such as a log odds-ratio, log risk-ratio, or difference in means. Each study also provides a standard error for this estimate,  $\sigma_i$ , which we assume is known, as is common in meta-analysis (although in practice, it will have been estimated from the data in that study). Let us start from the simplest model:

- *Fixed-effects meta-analysis* assumes that there is a single true effect size,  $\theta$ , so that

$$y_i \sim N(\theta, \sigma_i^2)$$

or equivalently,

$$y_i = \theta + \epsilon_i, \quad \text{where } \epsilon_i \sim N(0, \sigma_i^2)$$

- *Random-effects meta-analysis* allows the true effects,  $\theta_i$ , to vary between studies by assuming that they have a normal distribution around a mean effect,  $\theta$ :

$$y_i | \theta_i \sim N(\theta_i, \sigma_i^2), \quad \text{where } \theta_i \sim N(\theta, \tau^2)$$

So

$$y_i \sim N(\theta, \sigma_i^2 + \tau^2)$$

or equivalently,

$$y_i = \theta + u_i + \epsilon_i, \quad \text{where } u_i \sim N(0, \tau^2) \text{ and } \epsilon_i \sim N(0, \sigma_i^2)$$

Here  $\tau^2$  is the between-study variance and must be estimated from the data.

- *Fixed-effects meta-regression* extends fixed-effects meta-analysis by replacing the mean,  $\theta$ , with a linear predictor,  $\mathbf{x}_i\boldsymbol{\beta}$ :

$$y_i \sim N(\theta_i, \sigma_i^2), \quad \text{where } \theta_i = \mathbf{x}_i\boldsymbol{\beta}$$

or equivalently,

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i, \quad \text{where } \epsilon_i \sim N(0, \sigma_i^2)$$

Here  $\boldsymbol{\beta}$  is a  $k \times 1$  vector of coefficients (including a constant if fitted), and  $\mathbf{x}_i$  is a  $1 \times k$  vector of covariate values in study  $i$  (including a 1 if a constant is fit).

- *Random-effects meta-regression* allows for such residual heterogeneity (between-study variance not explained by the covariates) by assuming that the true effects follow a normal distribution around the linear predictor:

$$y_i | \theta_i \sim N(\theta_i, \sigma_i^2), \quad \text{where } \theta_i \sim N(\mathbf{x}_i\boldsymbol{\beta}, \tau^2)$$

so

$$y_i \sim N(\mathbf{x}_i\boldsymbol{\beta}, \sigma_i^2 + \tau^2)$$

or equivalently,

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i + \epsilon_i, \quad \text{where } u_i \sim N(0, \tau^2) \text{ and } \epsilon_i \sim N(0, \sigma_i^2)$$

Random-effects meta-regression can be considered either an extension to fixed-effects meta-regression that allows for residual heterogeneity or an extension to random-effects meta-analysis that includes study-level covariates.

Table 1 summarizes the relationships between these models and gives the corresponding Stata commands, which are summarized in the next section.

(Continued on next page)

Table 1. Summary of `metareg` and related Stata commands

	No covariates	With covariate(s)
Fixed-effects model	fixed-effects meta-analysis  <code>metan</code> with <code>fixedi</code> , <code>peto</code> , or no options	fixed-effects meta-regression (not recommended)  <code>vwls</code>
Random-effects model	random-effects meta-analysis  <code>metan</code> with <code>random</code> or <code>randomi</code> options	random-effects meta-regression (mixed-effects meta-regression)  <code>metareg</code>

### 3 Relation to other Stata commands

Both fixed- and random-effects meta-analysis are available in the user-written package `metan` (Harris et al. 2008). Random-effects meta-analysis can also be performed with `metareg` by not including any covariates (the method-of-moments estimate for between-study variance must be specified to produce identical results to the `metan` command). `metan` can also be used to generate the variables required by `metareg` containing the effect estimate and its standard error for each study from data in various other forms (Harris et al. 2008).

Fixed-effects meta-regression can be fit by weighted least squares by using the official Stata command `vwls` (see [R] `vwls`) with the weights  $1/\sigma_i^2$ . Fixed-effects meta-regression is not usually recommended, however, because it assumes that all the heterogeneity can be explained by the covariates, and it leads to excessive type I errors when there is residual, or unexplained, heterogeneity (Higgins and Thompson 2004; Thompson and Sharp 1999).

Random-effects meta-regression is closely related to the seldom-used “between-effects” model available in the official Stata command `xtreg` (see [XT] `xtreg`), with studies corresponding to units. Whereas meta-regression assumes that the within-study data have been summarized by an effect estimate,  $y_i$ , and its standard error,  $\sigma_i$ , for each study, `xtreg` requires data on individual observations, e.g., individual patient data. Meta-regression is often used on binary outcomes summarized by log odds-ratios or log risk-ratios and their standard errors, whereas `xtreg` is appropriate only for continuous outcomes. `xtreg` also uses different estimators from those available in `metareg`, which are outlined in section 5.1.

## 4 Background to examples

Our first example is from a meta-analysis of 28 randomized controlled trials of cholesterol-lowering interventions for reducing risk of ischemic heart disease (IHD). The outcome event was death from IHD or nonfatal myocardial infarction. These data are taken from table 1 of [Thompson and Sharp \(1999\)](#). Data from 25 of these trials were also published in [Thompson \(1993\)](#). The measure of effect size is the odds ratio, but statistical analysis is conducted on its natural logarithm, the log odds-ratio, because this has a sampling distribution more closely approximated by a normal distribution. The interventions are varied, with 18 trials of several different drugs, 9 trials of dietary interventions, and 1 trial of a surgical intervention. The eligibility criteria also differed—19 studies recruited only participants without known IHD on entry, 6 recruited only those with IHD, and 3 included those with or without IHD. The reduction in cholesterol varied among trials, as quantified by the difference in mean serum cholesterol concentrations between the treated and control subjects at the end of each trial. Interest focuses on estimating the odds ratio for any given degree of cholesterol reduction (e.g., 1 mmol/L), assuming that any effect on IHD is mediated through the reduction in serum cholesterol. The Stata dataset is named `cholesterol.dta`.

The second example is drawn from a systematic review of 10 randomized controlled trials of exercise as an intervention in the management of depression ([Lawlor and Hopker 2001](#)). Here the outcome, severity of depression, was measured on one of two numerical scales, and the measure of effect size was the standardized mean difference. There was considerable between-study heterogeneity in the results of the trials, and the authors considered eight study-level covariates that might explain this heterogeneity. We will focus on the five covariates selected by [Higgins and Thompson \(2004\)](#). The Stata dataset is named `xrcise4deprsn.dta`.

## 5 New and enhanced features

We now give details of each of the new and enhanced features available in this revision of `metareg`, as listed in section 1. Sections 5.1–5.3 are relevant to all uses of `metareg`. When there is a single continuous covariate, the fitted model can be presented graphically, as shown in section 5.4. Section 5.5 explores a permutation-based approach to calculating  $p$ -values, suggested by [Higgins and Thompson \(2004\)](#), who recommended its use when there are few studies and as a way of adjusting for multiple testing when there is more than one covariate of interest. Section 5.6 is intended for more advanced users only; it describes the postestimation facilities available after a `metareg` model has been fit, and it assumes some familiarity with random-effects models, as well as with Stata's graphics commands and postestimation tools.

### 5.1 Algorithm for REML estimation of $\tau^2$

All algorithms for random-effects meta-regression first estimate the between-study variance,  $\tau^2$ , and then estimate the coefficients,  $\beta$ , by weighted least squares by using the

weights  $1/(\sigma_i^2 + \tau^2)$ , where  $\sigma_i^2$  is the standard error of the estimated effect in study  $i$ . The default algorithm in `metareg` is REML, as advocated by [Thompson and Sharp \(1999\)](#).

The algorithm for REML estimation has been improved in this update of `metareg`. The original version used an iterative algorithm ([Morris 1983](#)) that was not guaranteed to converge and was only an approximation when the within-study standard errors varied. The original version of `metareg` sometimes misleadingly reported an estimate of  $\hat{\tau}^2 = 0$  when the algorithm was in fact diverging (for example, with the `cholesterol` data). This revised version of `metareg` instead directly maximizes the residual (restricted) log likelihood by using Stata's robust and well-tested `ml` command, avoiding the approximations and convergence problems of the previous method.

We decided not to implement the standard maximum likelihood (ML) estimator in this updated version of `metareg`. (To ensure all do-files written for the original version of `metareg` continue to work, however, the code of the original program is included in this package so that a request for the ML estimator can be handled by calling the original code.) Both REML and ML are iterative methods. Unlike REML, however, ML does not account for the degrees of freedom used in estimating the fixed effects. This can make a particular difference in meta-regression because the number of observations (studies) is often small. As a result, the ML estimate of  $\tau^2$  is often biased downward, leading to underestimated standard errors and anticonservative inference ([Thompson and Sharp 1999](#); [Sidik and Jonkman 2007](#)).

Further details of the methods for the estimation of  $\tau^2$  are given in section 7.1.

## 5.2 Knapp–Hartung variance estimator and associated t test

[Knapp and Hartung \(2003\)](#) introduced a novel estimator for the variances of the effect estimates in meta-regression. Their variance estimator amounts to calculating a quadratic form,  $q$ , and multiplying the usual variance estimates by  $q$  if  $q > 1$ . This estimator should be used with a  $t$  distribution when calculating  $p$ -values and confidence intervals. They found this procedure to have much more appropriate false-positive rates than the standard approach, a finding confirmed by [Higgins and Thompson \(2004\)](#) in more extensive simulations.

We therefore recommend this variance estimator and have made it the default in `metareg`. It is particularly suitable for estimation of standard errors and confidence intervals. However, it can be unreasonably conservative (false-positive rates below the nominal level) when the number of studies is particularly small, further reducing the already limited power. When there are few studies, the permutation test detailed in section 5.5 below has the potential to provide a better, though more computationally intensive, method for calculating  $p$ -values.

### 5.3 Enhancements to the output

The following additions have been made to the output of `metareg` that is displayed above the coefficient table:

- A measure of the percentage of the residual variation that is attributable to between-study heterogeneity ( $I^2_{\text{res}}$ )
- The proportion of between-study variance explained by the covariates (a type of adjusted  $R^2$  statistic)
- An overall test of all the covariates in the random-effects model

The iteration log is no longer displayed by default.

We will illustrate these additions by using the output of `metareg` in the simplest situation where a single continuous covariate is fit, using the `cholesterol` data as an example:

```
. use cholesterol
(Serum cholesterol reduction & IHD)
. metareg logor cholreduc, wsse(selogor)
```

Meta-regression	Number of obs =	28
REML estimate of between-study variance	tau2 =	.0097
% residual variation attributable to heterogeneity	I-squared_res =	31.34%
Proportion of between-study variance explained	Adj R-squared =	69.02%
With Knapp-Hartung modification		

	logor	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cholreduc		-.5056849	.1834858	-2.76	0.011	-.8828453 -.1285244
_cons		.1467225	.1374629	1.07	0.296	-.1358367 .4292816

#### Residual heterogeneity of the fixed-effects model

The residual heterogeneity statistic is the weighted sum of squares of the residuals from the fixed-effects meta-regression model and is a generalization of Cochran's  $Q$  from meta-analysis to meta-regression. To distinguish it from the total heterogeneity statistic  $Q$  that would be obtained from ordinary meta-analysis, i.e., without fitting any covariates, we will denote it by  $Q_{\text{res}}$  (Lipsey and Wilson [2001] denote the same statistic by  $Q_E$ ). A test of the null hypothesis of no residual (unexplained) heterogeneity can be obtained by comparing  $Q_{\text{res}}$  to a  $\chi^2$  distribution with  $n - k$  degrees of freedom. However, it is often more useful to quantify heterogeneity than to test for it (Higgins et al. 2003): The proportion of residual between-study variation due to heterogeneity, as opposed to sampling variability, is calculated as  $I^2_{\text{res}} = \max[0, \{Q_{\text{res}} - (n - k)\}/Q_{\text{res}}]$ , an obvious extension to the  $I^2$  measure in meta-analysis (Higgins et al. 2003).

From the value of  $I^2_{\text{res}}$  in the output above, 31% of the residual variation is due to heterogeneity, with the other 69% attributable to within-study sampling variability.

### Adjusted $R^2$

The proportion of between-study variance explained by the covariates can be calculated by comparing the estimated between-study variance,  $\hat{\tau}^2$ , with its value when no covariates are fit,  $\hat{\tau}_0^2$ . Adjusted  $R^2$  is the relative reduction in the between-study variance,  $R_{\text{adj}}^2 = (\hat{\tau}_0^2 - \hat{\tau}^2)/\hat{\tau}_0^2$ . It is possible for this to be negative if the covariates explain less of the heterogeneity than would be expected by chance, but the same is true for adjusted  $R^2$  in ordinary linear regression. It may be more common in meta-regression because the number of studies is often small.

In the above example, 69% of the between-study variance is explained by the covariate `cholreduc`, and the remaining between-study variance appears small at 0.0097. (It is coincidence that the figure of 69% also appears in the preceding subsection.)

### Joint test for all covariates

When more than one covariate is fit, `metareg` reports a test of the null hypothesis that the coefficients of the covariates are all zero, obtained from a multiparameter Wald test by using Stata's `test` command (see [R] `test`). The test statistic is compared to the appropriate  $F$  distribution if the default Knapp–Hartung adjustment is used. If `metareg`'s `z` option is used to specify the use of conventional variance estimates and tests for the effect estimates, a  $\chi^2$  distribution is used for the joint test. To simplify the output, this test is not displayed when only a single covariate is fit because it would give an identical  $p$ -value to the one displayed for the covariate in the regression table.

This gives one way of controlling the risk of false-positive findings when performing meta-regression with multiple covariates: we can use the overall model  $p$ -value to assess if there is evidence for an association of *any* of the covariates with the outcome. However, when a small  $p$ -value indicates that there is such evidence, it becomes harder to decide which, and how many, of the covariates there is good evidence for. Another method of dealing with this multiplicity issue that may help overcome this problem, though at the expense of longer computation time, is given in section 5.5 below.



► **Example**

We illustrate this joint test by using all five covariates available in the data on exercise for depression:

```
. use xrcise4deprsn
(Exercise for depression)
. metareg smd abstract-phd, wsse(sesmd)
```

Meta-regression	Number of obs	=	10
REML estimate of between-study variance	tau2	=	0
% residual variation attributable to heterogeneity	I-squared_res	=	0.00%
Proportion of between-study variance explained	Adj R-squared	=	100.00%
Simultaneous test for all covariates	Model F(5,4)	=	6.57
With Knapp-Hartung modification	Prob > F	=	0.0460

smd	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
abstract	-1.33993	.3892562	-3.44	0.026	-2.420678	-.2591814
duration	.1567629	.0616404	2.54	0.064	-.0143784	.3279041
itt	.4611682	.3883635	1.19	0.301	-.6171018	1.539438
alloc	-.4063866	.3503447	-1.16	0.311	-1.379099	.5663263
phd	-.0138045	.440595	-0.03	0.977	-1.237092	1.209483
_cons	-2.07241	.5683944	-3.65	0.022	-3.650526	-.4942942

Here  $\hat{\tau}^2$  is zero, and it follows that  $I_{\text{res}}^2 = 0\%$  and  $R_{\text{adj}}^2 = 100\%$ . The joint test for all five covariates gives a  $p$ -value of 0.046, indicating some evidence for an association of at least one of the covariates with the size of the treatment effect.

◀

## 5.4 Graph of the fitted model

When a single continuous covariate is fit, one common way to present the fitted model, sometimes referred to as a “bubble plot”, is to graph the fitted regression line together with circles representing the estimates from each study, sized according to the precision of each estimate (the inverse of its within-study variance,  $\sigma_i^2$ ). The **graph** option to **metareg** gives an easy way to produce such a plot, as illustrated in figure 1 for the cholesterol data.

```
. use cholesterol
(Serum cholesterol reduction & IHD)
. metareg logor cholreduc, wsse(selogor) graph
(output omitted)
```

(Continued on next page)

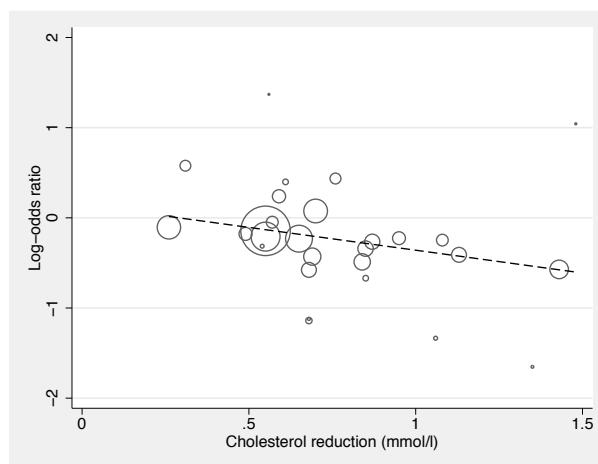


Figure 1. “Bubble plot” with fitted meta-regression line

An additional option, `randomsize`, is provided for those who prefer the size of the circles to depend on the weight of the study in the fitted random-effects meta-regression model (the inverse of its total variance,  $\sigma_i^2 + \hat{\tau}^2$ ). This makes only a slight difference to the example above because the estimated between-study variance,  $\hat{\tau}^2$ , is small; in general, though, it will give circles that vary less in size.

Those wishing to further customize the plot can use the `predict` command to generate fitted values followed by a `graph twoway` command (see section 5.6).

## 5.5 Permutation test

Higgins and Thompson (2004) proposed using a permutation test approach to calculating  $p$ -values in meta-regression. Permutation tests provide a nonparametric way of simulating data under the null hypothesis (see, e.g., Manly [2006]). Calculation of exact permutation  $p$ -values would be feasible when there are few studies by enumeration of all possible permutations, but for simplicity, we have implemented a permutation test based on Monte Carlo simulation, i.e., based on *random* permutations.

The algorithm is similar to other applications of permutation methods, and it is implemented with Stata’s `permute` command (see [R] `permute`). The covariates are randomly reallocated to the outcomes many times, and a  $t$  statistic is calculated each time. The true  $p$ -value for the relationship between a given covariate and the response is computed by counting the number of times these  $t$  statistics are greater than or equal to the observed  $t$  statistic. When multiple covariates are included in the meta-regression, the covariate values for a given study are kept together to preserve and account for their correlation structure. In meta-regression, unlike other regressions, the outcome consists of both the effect size and its standard error, and these must be kept together. This small complication makes it impossible to use `permute` on `metareg` directly from

the command line when there are multiple covariates, so we have written a `permute()` option for `metareg`. This option also implements the following extension, which adjusts  $p$ -values for multiple tests when there are several covariates.

### Multiplicity adjustment

When several covariates are used in meta-regression, either in several separate univariable meta-regressions or in one multiple meta-regression, there is an increased chance of at least one false-positive finding (type I error). The statistics obtained from the random permutations can be used to adjust for such multiple testing by comparing the observed  $t$  statistic for every covariate with the largest  $t$  statistic for any covariate in each random permutation. The proportion of times that the former equals or exceeds the latter gives the probability of observing a  $t$  statistic for any covariate as extreme or more extreme than that observed for a particular covariate, under the complete null hypothesis that all the regression coefficients are zero.

The number of random permutations must be specified—there is deliberately no default. We suggest that a small number (e.g., 100) be specified initially to check that the command is working as expected. The number should then be increased to at least 1,000, but 5,000 or 20,000 permutations may be necessary for sufficient precision ([Manly 2006](#); [Westfall and Young 1993](#)). Because the `permute()` option uses Stata's random-number generator, the `set seed` command should be used first if replicability of results is desired. When the `permute()` option is specified, the defaults are to use the method-of-moments estimate of  $\tau^2$  for reasons of speed and to not use the Knapp–Hartung modification to the standard errors.

By default, `permute()` performs multivariable meta-regression; i.e., all the covariates are entered into a single model in each permutation.

### ► Example

We illustrate the use of the `permute()` option by using the data on exercise for depression.

*(Continued on next page)*

```
. use xrcise4deprsn
(Exercise for depression)
. set seed 15160401
. metareg smd abstract-phd, wsse(sesmd) permute(20000)
Monte Carlo permutation test for meta-regression
Moment-based estimate of between-study variance
Without Knapp & Hartung modification to standard errors
P-values unadjusted and adjusted for multiple testing
```

	Number of obs =	10
	Permutations =	20000
smd	P	
	Unadjusted	Adjusted
abstract	0.023	0.089
duration	0.056	0.201
itt	0.311	0.721
alloc	0.313	0.736
phd	0.978	1.000

```
largest Monte Carlo SE(P) = 0.0033
```

```
WARNING:
```

```
Monte Carlo methods use random numbers, so results may differ between runs.
Ensure you specify enough permutations to obtain the desired precision.
```

The first column of the results table gives permutation  $p$ -values without an adjustment for multiplicity. The results are in good agreement with the  $p$ -values obtained in section 5.3 without using the permutation option but with the Knapp–Hartung modification. The second column gives  $p$ -values adjusted for multiplicity. We see that all the  $p$ -values are increased. After adjusting for multiple testing, there remains some weak evidence that results of studies published as an abstract differ on average from results of studies published as a full article. The adjusted  $p$ -value of 0.089 gives the probability under the complete null hypothesis (that all regression coefficients are zero) of a  $t$  statistic for *any* of the five covariates as extreme or more extreme as that observed for the covariate **abstract**. As [Higgins and Thompson \(2004\)](#) suggest, this can be interpreted as describing the degree of “surprise” one might have about the observed result for this covariate, considering that five covariates are being examined. This is less conservative than the Bonferroni adjusted  $p$ -value of  $0.0235 \times 5 = 0.1175$ .

The output also gives the largest Monte Carlo standard error of the calculated  $p$ -values as an indication of the degree of precision obtained by the specified number of random permutations. Standard errors and “exact” confidence intervals for each of the  $p$ -values can be obtained by using the **detail** suboption. (These can always be calculated afterward by using the **cii** command if this option was not specified.)

◀

#### □ Technical note

[Higgins and Thompson \(2004\)](#) originally proposed a slightly different permutation-based multiplicity adjustment: it compared the  $i$ th largest  $t$  statistic observed (for the

“ $i$ th most significant” covariate) with the  $i$ th largest  $t$  statistic in each random permutation. This adjustment was implemented in a revised version of `metareg` released previously on the Statistical Software Components archive. This adjustment has been found to be hard to interpret in practice, however, because for the second most significant covariate it effectively gives a *joint* test of the two covariates with the largest two observed  $t$  statistics (and similarly for third and subsequent covariates if more than two covariates are supplied). The resulting multiplicity-adjusted  $p$ -value can turn out to be either larger or smaller than the unadjusted  $p$ -value, which can appear counter-intuitive.

For this release of `metareg`, we have therefore chosen to implement a different permutation-based algorithm for multiplicity adjustment based on the one-step “maxT” method of [Westfall and Young \(1993\)](#). This adjustment compares the  $t$  statistic for every covariate with the largest  $t$  statistic in each random permutation. The resulting multiplicity-adjusted  $p$ -values are always as large as or (usually) larger than the unadjusted  $p$ -values. This procedure ensures weak control of the familywise error rate, defined as the probability that at least one null hypothesis is rejected when *all* the null hypotheses are true ([Shaffer 1995](#)). It does not guarantee strong control of the familywise error rate, however; i.e., when one or more null hypotheses are false, it does not guarantee control of the proportion of the remaining true null hypotheses that are incorrectly rejected, though such strong control should be achieved asymptotically as the number of studies increases ([Westfall and Young 1993](#); [Shaffer 1995](#)).

The false discovery rate ([Benjamini and Hochberg 1995](#)) and related procedures ([Newson and the ALSPAC Study Team 2003](#); [Storey, Taylor, and Siegmund 2004](#); [Wacholder et al. 2004](#)) have been suggested as an alternative method of multiplicity adjustment, but we have chosen not to implement such procedures in `metareg`. Such procedures are always either step-up or (more rarely) step-down algorithms. Although stepwise algorithms are suitable for hypothesis testing and often give greater power, the resulting adjusted  $p$ -values cannot be interpreted as giving the strength of evidence against the null hypothesis, the interpretation increasingly advocated in medicine and epidemiology ([Sterne and Davey Smith 2001](#)). In particular, stepwise methods may assign equal adjusted  $p$ -values to covariates with different unadjusted  $p$ -values.

□

### Suboptions to `permute()`

The `permute()` option can also be used to perform a set of single-variable meta-regressions at each permutation by adding the `univariable` suboption. This suboption reports permutation-based  $p$ -values for fitting a separate model for each covariate rather than including all the covariates in a multiple regression model. With several covariates, the execution time may be considerably longer than for multivariable meta-regression.

### ► Example

We add the `univariable` suboption to the previous example but reduce the number of permutations to cut down the computation time:

```
. metareg smd abstract-phd, wsse(sesmd) permute(5000, univariable)
Monte Carlo permutation test for single covariate meta-regressions
Moment-based estimate of between-study variance
Without Knapp & Hartung modification to standard errors
P-values unadjusted and adjusted for multiple testing
      Number of obs =      10
      Permutations  =    5000
```

smd	P	
	Unadjusted	Adjusted
abstract	0.021	0.043
duration	0.030	0.115
itt	0.384	0.946
alloc	0.330	0.861
phd	0.715	0.999

```
largest Monte Carlo SE(P) = 0.0069
```

**WARNING:**

Monte Carlo methods use random numbers, so results may differ between runs.  
Ensure you specify enough permutations to obtain the desired precision.

In these results, unlike those from the previous example, each covariate is fit in a separate model and so is not adjusted for the other covariates. The  $p$ -values do not differ greatly in this example, however.

◀

There is also a `joint()` suboption that requests a permutation  $p$ -value for a joint test of the variables specified. This can be particularly useful if a set of indicator variables is used to model a categorical covariate.

A joint test of covariates can be obtained without using a permutation approach by instead using the `test` or `testparm` (see [R] `test`) command after `metareg`.

A  $p$ -value for the joint test is not included in the multiplicity-adjustment procedure because the two are neither technically nor philosophically compatible.

► **Example**

We return to the `cholesterol` data, in which the `ihdentry` variable is a categorical covariate with three categories indicating whether the study included participants with known IHD on entry to the study, without known IHD, or both:

```
. use cholesterol
(Serum cholesterol reduction & IHD)
. tab ihdentry, gen(ihd)
```

Ischaemic heart disease on entry	Freq.	Percent	Cum.
Without known IHD	6	21.43	21.43
With IHD	19	67.86	89.29
With or without IHD	3	10.71	100.00
Total	28	100.00	

```
. metareg logor cholreduc ihd2 ihd3, wsse(selogor)
> permute(5000, joint(ihd2 ihd3))

Monte Carlo permutation test for meta-regression
Moment-based estimate of between-study variance
Without Knapp & Hartung modification to standard errors
joint1 : ihd2 ihd3
P-values unadjusted and adjusted for multiple testing
          Number of obs =      28
          Permutations  =    5000
```

logor	P	
	Unadjusted	Adjusted
cholreduc	0.009	0.028
ihd2	0.611	0.933
ihd3	0.907	0.999
joint1	0.883	

```
largest Monte Carlo SE(P) = 0.0069
```

**WARNING:**

Monte Carlo methods use random numbers, so results may differ between runs.  
Ensure you specify enough permutations to obtain the desired precision.

The  $p$ -value of 0.883 for the joint test of `ihd2` and `ihd3` indicates that there is very little evidence that the log odds-ratio differs among these three categories of studies, after adjusting for the degree of cholesterol reduction achieved in each study.



## 5.6 Postestimation tools for metareg

`metareg` is programmed as a Stata estimation command and so supports most of Stata's postestimation commands (except when the `permute()` option is used). (One deliberate exception is `lrtest`, which is not appropriate after `metareg` because the REML log likelihood cannot be used to compare models with different fixed effects, while the method of moments does not give a likelihood.)

Several quantities can be obtained by using `predict` after `metareg`, including fitted values and predicted random effects (empirical Bayes estimates). These can be useful for producing graphs of the fitted model and for model checking. Details of the syntax and options are given in sections 6.4 and 6.5, and section 7.4 contains the formulas used.

We now illustrate the use of some of the quantities available from `predict` in a graph. Using the exercise for `depression` data, we conduct a meta-regression of the standardized mean difference on the single covariate duration that describes the duration of follow-up in each study. Figure 2 shows the fitted line and the estimates from the separate studies that would be produced by the `graph` option to `metareg`, and it also includes the empirical Bayes estimates and shaded bands showing both confidence and prediction intervals (we would not recommend including all these features on a single graph in practice). It was produced by the following commands:

```
. use xrcise4deprsn, clear
(Exercise for depression)
. metareg smd duration, wsse(sesmd)
```

Meta-regression

REML estimate of between-study variance	Number of obs	=	10
% residual variation attributable to heterogeneity	tau2	=	.2019
Proportion of between-study variance explained	I-squared_res	=	55.83%
With Knapp-Hartung modification	Adj R-squared	=	55.16%

smd	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
duration	.2097633	.0802611	2.61	0.031	.0246808	.3948457
_cons	-2.907511	.7339255	-3.96	0.004	-4.599946	-1.215076

```
. predict fit
(option xb assumed; fitted values)
. predict stdp, stdp
. predict stdf, stdf
. predict xbu, xbu
. local t = invttail(e(df_r)-1, 0.025)
. gen confl = fit - `t'*stdp
. gen confu = fit + `t'*stdp
. gen predl = fit - `t'*stdf
. gen predu = fit + `t'*stdf
. sort duration
. twoway rarea predl predu duration || rarea confl confu duration
> || line fit duration
> || scatter smd duration [aw=1/sesmd^2], msymbol(Oh)
> || scatter xbu duration, msymbol(t)
> ||, legend(label(1 "Prediction interval") label(2 "Confidence interval")
> cols(1))
```



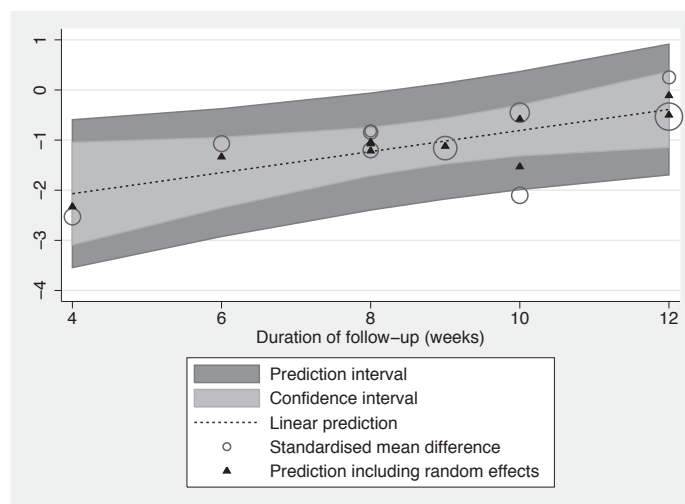


Figure 2. Confidence and prediction intervals and empirical Bayes estimates

The `stdp` option to `predict` gives the standard error of the fitted values excluding the random effects, commonly referred to as the standard error of the prediction. This standard error is used to draw a pointwise confidence interval, shown in light gray in figure 2, around the fitted line, illustrating our uncertainty about the position of the line. The `stdf` option to `predict` gives the standard deviation of the predicted distribution of the true value of the outcome in a future study with a given value of the covariate(s), commonly referred to as the standard error of the forecast. This standard error is used to draw a prediction interval, shown in dark gray in figure 2, around the fitted line, illustrating our uncertainty about the true effect we would predict in a future study with a known duration of follow-up. The prediction band will be wider than the confidence band unless  $\tau^2 = 0$ . The use of a  $t$  distribution in generating the intervals is an approximation, and opinions differ over the most appropriate degrees of freedom; we use  $n - k - 1$  here to be consistent with the  $n - 2$  used by Higgins, Thompson, and Spiegelhalter (Forthcoming) for confidence and prediction intervals in meta-analysis, where  $k = 1$ . The `xbu` option to `predict` gives the empirical Bayes estimates (predictions including random effects), shown as triangles in figure 2. These are our best estimates of the true effect in each study, assuming the fitted model is correct. If  $I_{\text{res}}^2$  is small, the empirical Bayes estimates will tend to lie well inside the prediction interval; if  $\tau^2 = 0$ , implying  $I_{\text{res}}^2 = 0$ , they will all lie on the fitted line.

The statistics available from `predict` can also be useful for model checking and checking for outliers and influential studies. This checking is best done graphically. One possibility is a normal probability plot of the standardized predicted random effects (equivalently, standardized empirical Bayes residuals, or standardized shrunk residuals; see figure 3). This probability plot can be used to check the assumption of normality of the random effects, although because this assumption has been used in

generating the predictions, only gross deviations are likely to be detected. Perhaps more usefully, the probability plot can be used to detect outliers:

```
. use cholesterol, clear
(Serum cholesterol reduction & IHD)
. qui metareg logor cholreduc, wsse(sellogor)
. capture drop usta
. predict usta, ustandard
. qnorm usta, mlabel(id)
```

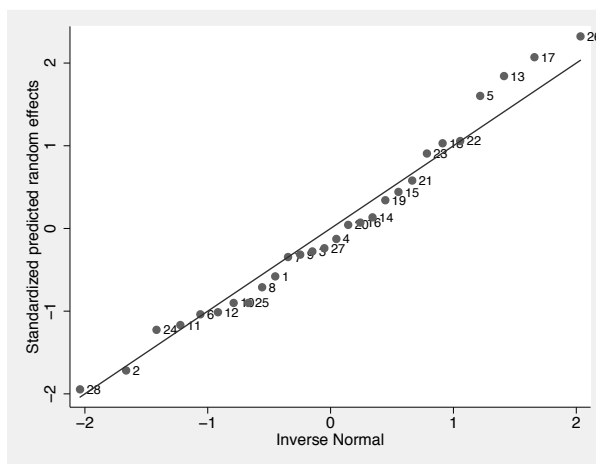


Figure 3. Normal probability plot of standardized shrunken residuals

Figure 3 suggests that the assumption of normal random effects is adequate, and there are no notable outliers because the largest standardized shrunken residual is only slightly over 2.

Other plots useful for model checking and identifying influential points in conventional linear regression may also be useful for meta-regression, for example, leverage-residual (L-R) plots, or plots of residuals versus either fitted values or a predictor; see [R] **regress postestimation** for further details of these and other plots (the various plot commands given there will not work after **metareg**, but it should be fairly straightforward to use **predict** followed by the appropriate **graph twoway** command to produce similar plots).

## 6 Syntax, options, and saved results

### 6.1 Syntax

The syntax of **metareg** has been revised somewhat from that of the original version (Sharp 1998). The original syntax should continue to work, but it is not documented

here. ML estimation of  $\tau^2$  is not supported by the updated `metareg` program, but if the old `bsest(ml)` option is used, the new program simply calls the original version, which is incorporated within the updated `metareg.ado` file.

```
metareg depvar [indepvars] [if] [in] wsse(varname) [, eform graph
    randomsize noconstant mm reml eb knapphartung z tau2test level(#)
    permute(#) [, univariable detail joint(varlist1 [ | varlist2 ...])] log
    maximize_options]
```

`by` can be used with `metareg`; see [D] `by`.

## 6.2 Options

`wsse(varname)` specifies the variable containing  $\sigma_i$ , the standard error of `depvar`, within each study. All values of `varname` must be greater than zero. `wsse()` is required.

`eform` indicates to output the exponentiated form of the coefficients and to suppress reporting of the constant. This option may be useful when `depvar` is the logarithm of a ratio measure, such as a log odds-ratio or a log risk-ratio.

`graph` requests a line graph of fitted values plotted against the first covariate in `indepvars`, together with the estimates from each study represented by circles. By default, the circle sizes depend on the precision of each estimate (the inverse of its within-study variance), which is the weight given to each study in the fixed-effects model.

`randomsize` is for use with the `graph` option. It specifies that the size of the circles will depend on the weights in the random-effects model rather than the precision of each estimate. These random-effects weights depend on the estimate of  $\tau^2$ .

*The remaining options will mainly be of interest to more advanced users:*

`noconstant` suppresses the constant term (intercept). This is rarely appropriate in meta-regression. *Note:* It might seem tempting to use the `noconstant` option in the cholesterol example to force the regression line through the origin, on the reasoning that an intervention that has no effect on cholesterol should have no effect on the odds of IHD. We would advise against using this option, however, both here and in most other circumstances. Using it here involves the assumption that the effect of the intervention on IHD is mediated entirely by cholesterol reduction. It also would not allow for measurement error in cholesterol reduction, which, through attenuation by errors (regression dilution bias), could lead to a nonzero intercept even when a zero intercept would be expected.

The `mm`, `reml`, and `eb` options are alternatives that specify the method of estimation of the additive (between-study) component of variance  $\tau^2$ :

**mm** specifies the use of method of moments to estimate the additive (between-study) component of variance  $\tau^2$ ; this is a generalization of the DerSimonian and Laird (1986) method commonly used for random-effects meta-analysis. For speed, this is the default when the **permute()** option is specified, because it is the only noniterative method.

**reml** specifies the use of REML to estimate the additive (between-study) component of variance  $\tau^2$ . This is the default unless the **permute()** option is specified. This revised version uses Stata's ML facilities to maximize the REML log likelihood. It will therefore not give identical results to the previous version of **metareg**, which used an approximate iterative method.

**eb** specifies the use of the “empirical Bayes” method to estimate  $\tau^2$  (Morris 1983).

**knapphartung** makes a modification to the variance of the estimated coefficients suggested by Knapp and Hartung (2003) and supported by Higgins and Thompson (2004), accompanied by the use of a  $t$  distribution in place of the standard normal distribution when calculating  $p$ -values and confidence intervals. This is the default unless the **permute()** option is specified.

**z** requests that the **knapphartung** modification not be applied and that the standard normal distribution be used to calculate  $p$ -values and confidence intervals. This is the default when the **permute()** option is specified with a fixed-effects model.

**tau2test** adds to the output two tests of  $\tau^2 = 0$ . The first is based on the residual heterogeneity statistic,  $Q_{\text{res}}$ . The second (not available if the **mm** option is also specified) is a likelihood-ratio test based on the REML log likelihood. These are two tests of the same null hypothesis (the fixed-effects model with  $\tau^2 = 0$ ), but the alternative hypotheses are different, as are the distributions of the test statistics under the null, so close agreement of the two tests is not guaranteed. Both tests are typically of little interest because it is more helpful to quantify heterogeneity than to test for it (see section 5.3).

**level(#)** specifies the confidence level, as a percentage, for confidence intervals. The default is **level(95)** or as set by **set level**; see [U] 20.7 Specifying the width of confidence intervals.

**permute(...)** calculates  $p$ -values by using a Monte Carlo permutation test. See section 6.3 below for more information about this option.

**log** requests the display of the iteration log during estimation of  $\tau^2$ . This is ignored if the **mm** option is specified, because this uses a noniterative method.

**maximize\_options** are ignored unless estimation of  $\tau^2$  is by REML. These options control the maximization process; see [R] **maximize**. They are ignored if the **mm** option is specified. You should never need to specify them; they are supported only in case problems in the REML estimation of  $\tau^2$  are ever reported or suspected.

### 6.3 Option for permutation test

The `permute()` option calculates  $p$ -values by using a Monte Carlo permutation test, as recommended by [Higgins and Thompson \(2004\)](#). To address multiple testing, `permute()` also calculates  $p$ -values for the most- to least-significant covariates, as the same authors also recommend.

The syntax of `permute()` is

```
permute(# [, univariable detail joint(varlist1 [ | varlist2 ... ]))
```

where `#` is required and specifies the number of random permutations to perform. Larger values give more precise  $p$ -values but take longer.

There are three suboptions:

**univariable** indicates that  $p$ -values should be calculated for a series of single covariate meta-regressions of each covariate in *varlist* separately, instead of a multiple meta-regression of all covariates in *varlist* simultaneously.

**detail** requests lengthier output in the format given by [R] **permute**.

`joint(varlist1 [ | varlist2 ... ])` specifies that a permutation  $p$ -value should also be computed for a joint test of the variables in each varlist.

The **eform**, **level()**, and **z** options have no effect when the `permute()` option is specified.

### 6.4 Syntax of predict

The syntax of **predict** (see [R] **predict**) following **metareg** is

```
predict [type] newvar [if] [in] [, statistic]
```

<i>statistic</i>	description
<b>xb</b>	fitted values; the default
<b>stdp</b>	standard error of the prediction
<b>stdf</b>	standard error of the forecast
<b>u</b>	predicted random effects
<b>ustandard</b>	standardized predicted random effects
<b>xbu</b>	prediction including random effects
<b>stdxbu</b>	standard error of <b>xbu</b>
<b>hat</b>	leverage (diagonal elements of hat matrix)

These statistics are available both in and out of sample; type **predict ... if e(sample)** ... if wanted only for the estimation sample.

## 6.5 Options for predict

**xb**, the default, calculates the linear prediction,  $\mathbf{x}_i\mathbf{b}$ , that is, the fitted values *excluding* the random effects.

**stdp** calculates the standard error of the prediction (the standard error of the fitted values excluding the random effects).

**stdf** calculates the standard error of the forecast. This gives the standard deviation of the predicted distribution of the true value of *depvar* in a future study, with the covariates given by *varlist*.  $\text{stdf}^2 = \text{stdp}^2 + \hat{\tau}^2$ .

**u** calculates the predicted random effects,  $u_i$ . These are the best linear unbiased predictions of the random effects, also known as the empirical Bayes (or posterior mean) estimates of the random effects, or as shrunken residuals.

**ustandard** calculates the standardized predicted random effects, i.e., the predicted random effects,  $u_i$ , divided by their (unconditional) standard errors. These may be useful for diagnostics and model checking.

**xbu** calculates the prediction *including* the random effects,  $\mathbf{x}_i\mathbf{b} + u_i$ , also known as the empirical Bayes estimates of the effects for each study.

**stdxbu** calculates the standard error of the prediction including random effects.

**hat** calculates the leverages (the diagonal elements of the projection hat matrix).

## 6.6 Saved results

When the `permute()` option is not specified, **metareg** saves the following in `e()`:

### Scalars

<code>e(N)</code>	number of observations	<code>e(tau2)</code>	estimate of $\tau^2$
<code>e(df_m)</code>	model degrees of freedom	<code>e(Q)</code>	Cochran's $Q$
<code>e(df_Q)</code>	degrees of freedom for test of $Q = 0$	<code>e(I2)</code>	$I$ -squared
<code>e(df_r)</code>	residual degrees of freedom (if $t$ tests used)	<code>e(q_KH)</code>	Knapp–Hartung variance modification factor
<code>e(remll)</code>	REML log likelihood	<code>e(remll_c)</code>	REML log likelihood, comparison model
<code>e(chi2_c)</code>	$\chi^2$ for comparison test	<code>e(tau2_0)</code>	$\tau^2$ , constant-only model
<code>e(F)</code>	model $F$ statistic	<code>e(chi2)</code>	model $\chi^2$

### Macros

<code>e(cmd)</code>	<b>metareg</b>	<code>e(depvar)</code>	name of dependent variable
<code>e(predict)</code>	program used to implement <b>predict</b>	<code>e(method)</code>	REML, Method of moments, or Empirical Bayes
<code>e(wsse)</code>	name of <code>wsse()</code> variable	<code>e(properties)</code>	<b>b</b> <b>V</b>

### Matrices

<code>e(b)</code>	coefficient vector	<code>e(V)</code>	variance–covariance matrix of estimators
-------------------	--------------------	-------------------	--

### Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

`metareg`, `permute()` saves the following in `r()`:

Scalars			
<code>r(N)</code>	number of observations		
Matrices			
<code>r(b)</code>	observed $t$ statistics, $T_{\text{obs}}$	<code>r(p)</code>	observed proportions
<code>r(c)</code>	count when $ T  \geq  T_{\text{obs}} $	<code>r(reps)</code>	number of nonmissing results

## 7 Methods and formulas

The residual heterogeneity statistic,  $Q_{\text{res}}$ , is the residual weighted sum of squares from the fixed-effects model and is the same as the goodness-of-fit statistic computed by `vls`:

$$Q_{\text{res}} = \sum_i \left( \frac{y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}}{\sigma_i} \right)^2$$

The proportion of residual variation due to heterogeneity is

$$I^2 = \max \left\{ \frac{Q_{\text{res}} - (n - k)}{Q_{\text{res}}}, 0 \right\}$$

The proportion of the between-study variance explained by the covariates (adjusted  $R$ -squared) is  $R_a^2 = (\hat{\tau}_0^2 - \hat{\tau}^2) / \hat{\tau}_0^2$ , where  $\hat{\tau}^2$  and  $\hat{\tau}_0^2$  are the estimates of the between-study variance in models with and without the covariates, respectively.

### 7.1 Estimation of $\tau^2$

Several different algorithms have been proposed for estimation of the between-study variance,  $\tau^2$ , in meta-analysis ([Sidik and Jonkman 2007](#)) and meta-regression ([Thompson and Sharp 1999](#)). Three algorithms are available in this version of `metareg`. In each case, if the estimated value of  $\tau^2$  is negative, it is set to zero.

Method of moments is the only noniterative method, so it has the advantages of speed and robustness. It is the natural extension of the [DerSimonian and Laird \(1986\)](#) estimate commonly used in random-effects meta-analysis. The method-of-moments estimate of  $\tau^2$  is obtained by equating the observed value of  $Q_{\text{res}}$  to its expected value under the random-effects model, giving ([DuMouchel and Harris 1983](#), eq. 3.12)

$$\hat{\tau}_{\text{MM}}^2 = \frac{Q_{\text{res}} - (n + k)}{\sum_i \{1/\sigma_i^2(1 - h_i)\}}$$

Here  $h_i$  is the  $i$ th diagonal element of the hat matrix  $\mathbf{X}(\mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{V}_0^{-1}$ , where  $\mathbf{V}_0 = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ .

The iterative methods below use  $\hat{\tau}_{\text{MM}}^2$  as a starting value (this is a change from the original version of `metareg` ([Sharp 1998](#)), which used zero as a starting value).

REML estimation of  $\tau^2$  is based on maximization of the residual (or restricted) log likelihood,

$$L_R(\tau^2) = -\frac{1}{2} \sum_i \left\{ \log(\sigma_i^2 + \tau^2) + \frac{(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})^2}{\sigma_i^2 + \tau^2} \right\} - \frac{1}{2} \log |\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}|$$

where  $\mathbf{V} = \text{diag}(\sigma_1^2 + \tau^2, \sigma_2^2 + \tau^2, \dots, \sigma_n^2 + \tau^2)$  and  $\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}$  (Harville 1977). This log likelihood is maximized by Stata's `ml` command, using the `d0` method, which calculates all derivatives numerically.

The “empirical Bayes” estimator of  $\tau^2$  is so named because of its introduction in an article on empirical Bayes inference by Morris (1983), although as he states, any approximately unbiased estimate of  $\tau^2$  could be used in such a setting. Thompson and Sharp (1999) found it to give substantially larger estimates of  $\tau^2$  than other methods. Others suggest it performs well in simulations based on  $2 \times 2$  tables (Berkey et al. 1995; Sidik and Jonkman 2007), although this may be due to overestimation of the within-study standard errors in small studies by the conventional (Woolf) estimate rather than the properties of the empirical Bayes method itself (Sutton and Higgins 2008). It can also be considered to be a method-of-moments estimator, formed by equating the weighted sum of squares of the residuals from the random-effects model to its expected value (Knapp and Hartung 2003). It is found by iterating the following equation (Morris 1983; Berkey et al. 1995):

$$\hat{\tau}_{\text{EB}}^2 = \frac{n/(n-k) \sum_i \left\{ (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})^2 / (\sigma_i^2 + \hat{\tau}_{\text{EB}}^2) - \sigma_i^2 \right\}}{\sum_i (\sigma_i^2 + \hat{\tau}_{\text{EB}}^2)^{-1}}$$

At each iteration,  $\hat{\boldsymbol{\beta}}$  must be reestimated using a weighted least-squares regression of  $\mathbf{y}$  on  $\mathbf{X}$  with the weights  $1/(\sigma_i^2 + \hat{\tau}_{\text{EB}}^2)$ .

## 7.2 Estimation of $\beta$

Once  $\tau^2$  has been estimated by one of the methods above, the estimated coefficients,  $\hat{\boldsymbol{\beta}}$ , are obtained by a weighted least-squares regression of  $\mathbf{y}$  on  $\mathbf{X}$  with the weights  $1/(\sigma_i^2 + \hat{\tau}^2)$ . The conventional estimate of the variance-covariance matrix of the coefficients is  $(\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1}$ , where  $\hat{\mathbf{V}} = \text{diag}(\sigma_1^2 + \hat{\tau}^2, \sigma_2^2 + \hat{\tau}^2, \dots, \sigma_n^2 + \hat{\tau}^2)$ .

## 7.3 Knapp–Hartung variance modification

Knapp and Hartung (2003) proposed multiplying the conventional estimate of the variance of the coefficients given above by  $\max(q, 1)$ , where the *Knapp–Hartung variance modification factor* is

$$q = \frac{1}{n-k} \sum_i \frac{(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})^2}{\sigma_i^2 + \hat{\tau}^2}$$

With the “empirical Bayes” estimator of  $\hat{\tau}^2$ ,  $q = 1$ , so this modification has no effect (Knapp and Hartung 2003).



## 7.4 Methods and formulas for predict

The *standard error of the prediction* (**stdp**) is  $s_{p_i} = \sqrt{\mathbf{x}_i(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{x}_i'}$ .

The *leverages*, or diagonal elements of the projection matrix (**hat**), are

$$h_i = s_{p_i}^2 / (\sigma_i^2 + \tau^2)$$

The *standard error of the forecast* (**stdf**) is  $s_{f_i} = \sqrt{s_{p_i}^2 + \tau^2}$ .

Denote the previously estimated coefficient vector by **b**, and let  $\lambda_i = \hat{\tau}^2 / (\sigma_i^2 + \hat{\tau}^2)$  denote the empirical Bayes shrinkage factor for the  $i$ th observation.

The *predicted random effects* (**u**) are  $u_i = \lambda_i(y_i - \mathbf{x}_i\mathbf{b})$ .

The *standardized predicted random effects* (**ustandard**) are

$$u_{s_j} = (y_i - \mathbf{x}_i\mathbf{b}) / \sqrt{\sigma_i^2 + \tau^2 - s_{p_i}^2}$$

The *prediction including random effects* (**xbu**), or empirical Bayes estimate, is

$$\mathbf{x}_i\mathbf{b} + u_i = \lambda_i y_i + (1 - \lambda_i)\mathbf{x}_i\mathbf{b}$$

The *standard error of the prediction including random effects* (**stdxbu**) is

$$\sqrt{\lambda_i^2(\sigma_i^2 + \tau^2) + (1 - \lambda_i^2)s_{p_i}^2}$$

## 8 Acknowledgments

Stephen Sharp gave permission to release this package under the same name as his original Stata package for meta-regression and to incorporate his code. Debbie Lawlor gave permission to use the example dataset on exercise for depression and provided additional unpublished data. We thank Simon Thompson for his helpful comments on the manuscript, and we thank the organizers of and participants at a meeting in Park City, Utah, in 2005 for discussions that influenced the output displayed by **metareg**. Finally, we wish to thank the referee for helpful comments, which led to improvements in the program and the article.

## 9 References

- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 57: 289–300.
- Berkey, C. S., D. C. Hoaglin, F. Mosteller, and G. A. Colditz. 1995. A random-effects regression model for meta-analysis. *Statistics in Medicine* 14: 395–411.

- Davey Smith, G., M. Egger, and A. N. Phillips. 1997. Meta-analysis: Beyond the grand mean? *British Medical Journal* 315: 1610–1614.
- DerSimonian, R., and N. Laird. 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials* 7: 177–188.
- DuMouchel, W. H., and J. E. Harris. 1983. Bayes methods for combining the results of cancer studies in humans and other species. *Journal of the American Statistical Association* 78: 293–308.
- Harris, R. J., M. J. Bradburn, J. J. Deeks, R. M. Harbord, D. G. Altman, and J. A. C. Sterne. 2008. metan: fixed- and random-effects meta-analysis. *Stata Journal* 8: 3–28.
- Harville, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72: 320–338.
- Higgins, J. P. T., and S. G. Thompson. 2004. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* 23: 1663–1682.
- Higgins, J. P. T., S. G. Thompson, J. J. Deeks, and D. G. Altman. 2002. Statistical heterogeneity in systematic reviews of clinical trials: A critical appraisal of guidelines and practice. *Journal of Health Services Research and Policy* 7: 51–61.
- . 2003. Measuring inconsistency in meta-analyses. *British Medical Journal* 327: 557–560.
- Higgins, J. P. T., S. G. Thompson, and D. J. Spiegelhalter. Forthcoming. A reevaluation of random-effects meta-analysis. *Journal of the Royal Statistics Society, Series A (Statistics in Society)*.
- Knapp, G., and J. Hartung. 2003. Improved tests for a random-effects meta-regression with a single covariate. *Statistics in Medicine* 22: 2693–2710.
- Lawlor, D. A., and S. W. Hopker. 2001. The effectiveness of exercise as an intervention in the management of depression: Systematic review and meta-regression analysis of randomised controlled trials. *British Medical Journal* 322: 763.
- Lipsey, M. W., and D. B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- Manly, B. F. J. 2006. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC.
- Morris, C. N. 1983. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association* 78: 47–55.
- Newson, R., and the ALSPAC Study Team. 2003. Multiple-test procedures and smile plots. *Stata Journal* 3: 109–132.
- Shaffer, J. P. 1995. Multiple hypothesis testing. *Annual Review of Psychology* 46: 561–584.

- Sharp, S. 1998. sbe23: Meta-analysis regression. *Stata Technical Bulletin* 42: 16–22. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, pp. 148–155. College Station, TX: Stata Press.
- Sidik, K., and J. N. Jonkman. 2007. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* 26: 1964–1981.
- Sterne, J. A. C., and G. Davey Smith. 2001. Sifting the evidence—what’s wrong with significance tests? *British Medical Journal* 322: 226–231.
- Storey, J. D., J. E. Taylor, and D. Siegmund. 2004. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 66: 187–205.
- Sutton, A. J., and J. P. T. Higgins. 2008. Recent developments in meta-analysis. *Statistics in Medicine* 27: 625–650.
- Thompson, S. G. 1993. Controversies in meta-analysis: The case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research* 2: 173–192.
- Thompson, S. G., and J. P. T. Higgins. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 21: 1559–1573.
- . 2005. Can meta-analysis help target interventions at individuals most likely to benefit? *Lancet* 365: 341–346.
- Thompson, S. G., and S. J. Sharp. 1999. Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine* 18: 2693–2708.
- Wacholder, S., S. Chanock, M. Garcia-Closas, L. El ghormli, and N. Rothman. 2004. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *JNCI Cancer Spectrum* 96: 434–442.
- Westfall, P. H., and S. S. Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: Wiley.

#### About the authors

Roger Harbord is a research associate in medical statistics in the Department of Social Medicine at the University of Bristol, UK. He is a co-convenor of the Cochrane Collaboration’s Screening and Diagnostic Tests Methods Group.

Julian Higgins is a senior statistician in the MRC Biostatistics Unit at the University of Cambridge, UK. He is an honorary visiting fellow of the UK Cochrane Centre in Oxford; an editor of the *Cochrane Handbook for Systematic Reviews of Interventions*, published by Wiley; and a coauthor of the book *Introduction to Meta-Analysis*, published by Wiley.