**Description**

`bmagraph coefdensity` provides posterior density plots for regression coefficients after the `bmaregress` command. This command is useful for exploratory analysis of regression coefficients. For a given predictor, `bmagraph coefdensity` shows the posterior probability of that predictor not being included in a model and, separately, the smooth density of its coefficient conditional on the predictor being included.

**Quick start**

Plot the analytical posterior density of the coefficient for predictor x after a Bayesian model averaging (BMA) regression model with a fixed g

```
   bmagraph coefdensity {x}
```

Same as above, but plot the MCMC-sample posterior density instead of the analytical one; `bmacoefsample` must be run first

```
   bmaregress, saving(bamodelmcmc)
   bmacoefsample
   bmagraph coefdensity {x}, mcmcsample
```

Plot posterior densities of coefficients for predictors x1 and x2, and customize the look of the vertical probability-mass line at zero for the second graph

```
   bmagraph coefdensity {x1 x2}, przeroline2opts(lcolor(green))
```

Same as above, but do not show the lines for probability mass at zero

```
   bmagraph coefdensity {x1 x2}, noprzerolines
```

Customize the kernel function used and the line pattern for the graph for \{x1\}

```
   bmagraph coefdensity {x1 x2}, dens1opts(kernel(gaussian) lpattern(dot))
```

Plot densities for all coefficients on one graph in two rows

```
   bmagraph coefdensity _all, combine(rows(2))
```

Plot the MCMC-sample posterior density of the coefficient for predictor x after a BMA regression model with a random g

```
   bmaregress, saving(bamodelfmcmc2)
   bmacoefsample
   bmagraph coefdensity {x}
```

Equivalent to the above `bmagraph coefdensity`

```
   bmagraph coefdensity {x}, mcmcsample
```
Menu

Statistics > Bayesian model averaging > Coefficient densities

Syntax

Density plot for one coefficient

\begin{verbatim}
bmagraph coefdensity \{coef\} [, singleopts]
\end{verbatim}

Density plot for multiple coefficients

\begin{verbatim}
bmagraph coefdensity coefspec [, multiopts]
\end{verbatim}

Density plot for all coefficients

\begin{verbatim}
bmagraph coefdensity _all [, multiopts]
\end{verbatim}

coef is an unabbreviated name of the variable or a factor level used in the bmaregress model. coefspec may be a regression coefficient \{coef\}, a list of coefficients \{coef1\}, \{coef2\}, etc., or, equivalently, \{coef1 coef2 \ldots\}. _all refers to all regression coefficients.

singleopts

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>analytic</td>
</tr>
<tr>
<td>mcmcsample</td>
</tr>
<tr>
<td>name(name, ...)</td>
</tr>
<tr>
<td>saving(filename, ...)</td>
</tr>
<tr>
<td>[no] przeroline</td>
</tr>
<tr>
<td>przerolineopt(line_options)</td>
</tr>
</tbody>
</table>

Density options
densopts

<table>
<thead>
<tr>
<th>Density options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y axis, X axis, Titles, Legend, Overall</td>
</tr>
<tr>
<td>twoway_options</td>
</tr>
</tbody>
</table>

any options other than by(), name(), and saving() documented in [G-3] \textit{twoway_options}
## multiopts

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>analytic</td>
<td>plot analytical posterior density; default for fixed $g$</td>
</tr>
<tr>
<td>mcmcsample</td>
<td>plot MCMC-sample posterior density; only choice for random $g$</td>
</tr>
<tr>
<td>name(namespec,...)</td>
<td>specify names of graphs</td>
</tr>
<tr>
<td>saving(filespec,...)</td>
<td>save graphs in files</td>
</tr>
<tr>
<td>[no]przerolines</td>
<td>plot or hide vertical lines for probability mass at zero</td>
</tr>
<tr>
<td>przerolineopts(line_options)</td>
<td>affect rendition of all vertical lines for probability mass at zero</td>
</tr>
<tr>
<td>przeroline#opts(line_options)</td>
<td>affect rendition of #th vertical line for probability mass at zero</td>
</tr>
<tr>
<td>combine[ (grcombineopts) ]</td>
<td>display plots on one graph</td>
</tr>
<tr>
<td>Density options</td>
<td></td>
</tr>
<tr>
<td>dens#opts(densopts)</td>
<td>density options for #th plot</td>
</tr>
<tr>
<td>densopts</td>
<td>density options for all plots</td>
</tr>
<tr>
<td>Y axis, X axis, Titles, Legend, Overall</td>
<td>any options other than by(), name(), and saving() documented in [G-3] twoway_options</td>
</tr>
<tr>
<td>twoway_options</td>
<td></td>
</tr>
</tbody>
</table>

### Options

- **analytic** specifies that the density be computed analytically. This is the default for models with a fixed $g$ and is not available for models with a random $g$. This option may not be specified together with option mcmcsample.

- **mcmcsample** specifies that the density be estimated from a Markov chain Monte Carlo (MCMC) sample of coefficients. This option uses an MCMC sample generated by the *bmacoefsample* command and may not be specified together with option analytic. This is the only choice for models with a random $g$; that is, option mcmcsample is implied with a random $g$. With a fixed $g$, if an MCMC sample is not available, option mcmcsample is not allowed. You must use *bmacoefsample* to generate an MCMC sample to produce coefficient density graphs based on sample estimates.

- **name(namespec[, replace])** specifies the name of the graph or multiple graphs. See [G-3] *name_option* for a single graph. If multiple graphs are produced, then the argument of name() is either a list of names or stub, in which case graphs are named stub1, stub2, and so on. With multiple graphs, if name() is not specified, name(Graph__#, replace) is assumed; thus, the produced graphs may be replaced by subsequent *bma* commands.

  The replace suboption causes existing graphs with the specified name or names to be replaced.

- **saving(filespec[, replace])** specifies the filename or filenames to use to save the graph or multiple graphs to disk. See [G-3] *saving_option* for a single graph. If multiple graphs are produced, then the argument of saving() is either a list of filenames or stub, in which case graphs are saved with filenames stub1, stub2, and so on.

  The replace suboption specifies that the file (or files) may be replaced if it already exists.

- **przerolines** and **noprzerolines** show or hide the vertical lines for probability mass at zero. By default, the lines are shown.
przerolineopts(\textit{line_options}) and \texttt{przeroline\#opts(line_options)} affect the rendition of the vertical lines for probability mass at zero; see \cite{G-3} \textit{line_options}. \texttt{przerolineopts()} controls the look of all vertical lines but may be overridden for specific lines by using the respective options \texttt{przeroline\#opts()}.\par

\texttt{combine[(grcombineopts)]} specifies the display of all plots of coefficients as subgraphs on one graph. By default, a separate graph is produced for each plot when multiple coefficients are specified.\par

\textit{grcombineopts} is any of the options documented in \cite{G-2} \texttt{graph combine}.\par

\textbf{Density options}\par

\texttt{dens\#opts(densopts)} specifies density options for the \#th density plot.\par

\textit{densopts} specify options for the (kernel) density plot; see the options documented in \cite{G-2} \texttt{graph twoway kdensity}, except options \texttt{horizontal} and \texttt{range(varname)}. When option \texttt{analytic} is assumed, the density functions are known; thus, the kernel density estimation is not needed. In that case, \textit{densopts} include only \texttt{n()} and \texttt{cline} options, described in \cite{G-2} \texttt{graph twoway kdensity}.\par

\textbf{Y axis, X axis, Titles, Legend, Overall }\par

\textit{twoway_options} are any of the options documented in \cite{G-3} \texttt{twoway_options}, excluding \texttt{by()}, \texttt{name()}, or \texttt{saving()}. These include, for instance, options for titling the graph (see \cite{G-3} \texttt{title_options}).\par

**Remarks and examples** \hfill \texttt{stata.com}\par

The \texttt{bmaregress} command computes and reports only the posterior means and standard deviations of the regression coefficients. But we can use \texttt{bmagraph coefdensity} to plot the entire posterior densities for the regression coefficients.\par

A posterior density for a regression coefficient in a BMA linear regression is a mixture of a discrete and continuous components. A discrete component corresponds to the posterior probability of noninclusion, the probability that the corresponding predictor is not included in a model after observing the data, which is equivalent to the posterior probability that the coefficient is equal to zero. This component is represented by a red vertical line with a probability mass at zero equal to one minus the posterior inclusion probability (PIP). A continuous component corresponds to the continuous density conditional on the predictor being included or, equivalently, on the coefficient not being equal to zero.\par

For a BMA linear regression with a fixed \texttt{g}, the posterior densities of regression coefficients are known and can be plotted at any prespecified \texttt{x}-axis points; see Posterior distributions of regression coefficients in \texttt{Methods and formulas} of \cite{BMA} \texttt{bmaregress}. We refer to these densities as analytical posterior densities. Alternatively, if an MCMC sample of regression coefficients, as produced by \texttt{bmacoefsample}, is available, the posterior densities can be estimated from this sample by using kernel density estimation (see \cite{R} \texttt{kdensity}) when you specify the \texttt{mcmcsample} option. We refer to these posterior densities as MCMC-sample posterior densities. With a random \texttt{g}, this is the only option because analytical posterior densities are not available. That is, with a random \texttt{g}, you must run the \texttt{bmacoefsample} first before you can run \texttt{bmagraph coefdensity}.\par


Example 1: Posterior density plots for BMA linear regression using enumeration

Recall the performance dataset (Chatterjee and Hadi 2012, sec. 3.3) analyzed in example 1 of [BMA] bmaregress, where the employees’ satisfaction with their supervisors, rating, is modeled by six potential predictors.

We fit a BMA linear regression for the rating outcome using all six other variables as potential predictors. By default, the model space is fully explored using enumeration.

```
use https://www.stata-press.com/data/r18/performance
data on employee satisfaction with supervisor
.bmaregress rating complaints-advance
Enumerating models ...
Computing model probabilities ...
Bayesian model averaging          No. of obs = 30
Linear regression                No. of predictors = 6
Model enumeration                Groups = 6
                                   Always = 0
Priors:                           No. of models = 64
Models: Beta-binomial(1, 1)      For CPMP >= .9 = 10
Cons.: Noninformative            Mean model size = 1.699
Coef.: Zellner's g               Shrinkage, g/(1+g) = 0.9730
   g: Benchmark, g = 36
sigma2: Noninformative           Mean sigma2 = 52.302

<table>
<thead>
<tr>
<th>rating</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Group</th>
<th>PIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>complaints</td>
<td>0.7052859</td>
<td>0.1224289</td>
<td>1</td>
<td>.99973</td>
</tr>
<tr>
<td>learning</td>
<td>0.0603014</td>
<td>0.1285281</td>
<td>3</td>
<td>.25249</td>
</tr>
<tr>
<td>advance</td>
<td>-0.0167921</td>
<td>0.073883</td>
<td>6</td>
<td>.13148</td>
</tr>
<tr>
<td>privileges</td>
<td>-0.0074174</td>
<td>0.0488635</td>
<td>2</td>
<td>.10998</td>
</tr>
<tr>
<td>raises</td>
<td>0.0070789</td>
<td>0.0670475</td>
<td>4</td>
<td>.10642</td>
</tr>
<tr>
<td>critical</td>
<td>0.0009713</td>
<td>0.0437848</td>
<td>5</td>
<td>.098534</td>
</tr>
<tr>
<td>_cons</td>
<td>14.8472</td>
<td>7.874219</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
```

Note: Coefficient posterior means and std. dev. estimated from 64 models.
Note: Default priors are used for models and parameter g.

For a fixed $g$, as in our example, the posterior distributions for the regression coefficients can be computed analytically, so we can use `bmagraph coefdensity` directly after `bmaregress`.

Let’s look at the distributions of regression coefficients for `complaints` and `learning`. Similarly to other Stata Bayesian commands, we use the curly-braces notation to refer to model parameters—here the regression coefficients.
Let us inspect the analytical marginal posterior density of the coefficient for `complaints`.

```
  . bmagraph coefdensity {complaints}
```

![Analytical posterior density graph](image)

For coefficient `complaints` (or using its full name `{rating:complaints}`), the probability of noninclusion is very low, 0.0003, so the red line that represents it is not even visible on the graph. Thus, the posterior density of `complaints` is essentially a continuous density, with a mean of roughly 0.7 and with a slightly heavier left tail. Most of the mass of the distribution is between roughly 0.025 and 1.1, away from 0.

Let’s plot the analytical marginal posterior density for the `learning` coefficient and request a cyan color for the density line:

```
  . bmagraph coefdensity {learning}, lcolor(cyan)
```

![Analytical posterior density graph](image)

The probability of a noninclusion for `learning` is about 0.75, so we see the red vertical line with the horizontal reference line at roughly 0.75. The look of the vertical line can be controlled via the `przerolineopts()` option. Conditional on the inclusion, the continuous density has its mass between roughly $-0.2$ and 0.7, which includes 0.

The posterior means and standard deviations reported by `bmaregress` are computed with respect to the above mixtures of distributions.
Instead of showing the density of each coefficient separately, we can plot the densities of all coefficients by specifying _all and request they be plotted on the same graph using the combine() option.

```
. bmggraph coefdensity _all, combine(rows(2))

Analytical posterior density
```

The rows(2) suboption of combine() specifies that plots be organized in two rows.
Example 2: Posterior density plots for BMA linear regression using sampling

We refit the same BMA model as in example 1 but now use the sampling option to request the MCMC model composition (MC3) algorithm instead of the default model enumeration.

```
.bmaregress rating complaints-advance, sampling rseed(18)
Burn-in ...
Simulation ...
Computing model probabilities ...
Bayesian model averaging
No. of obs = 30
Linear regression
No. of predictors = 6
MC3 sampling
Groups = 6
Always = 0
No. of models = 32
For CPMP >= .9 = 10
Priors:
Mean model size = 1.699
Models: Beta-binomial(1, 1)
Burn-in = 2,500
Cons.: Noninformative
MCMC sample size = 10,000
Coef.: Zellner's g
Acceptance rate = 0.2417
Shrinkage, g/(1+g) = 0.9730
sigma2: Noninformative
Mean sigma2 = 52.292
Sampling correlation = 0.9990
```

<table>
<thead>
<tr>
<th>rating</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Group</th>
<th>PIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>complaints</td>
<td>0.705479</td>
<td>0.1218881</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>learning</td>
<td>0.0601919</td>
<td>0.128269</td>
<td>3</td>
<td>.25234</td>
</tr>
<tr>
<td>advance</td>
<td>-0.0167514</td>
<td>0.0737415</td>
<td>6</td>
<td>.13141</td>
</tr>
<tr>
<td>privileges</td>
<td>-0.0074265</td>
<td>0.048844</td>
<td>2</td>
<td>.10996</td>
</tr>
<tr>
<td>raises</td>
<td>0.0069949</td>
<td>0.0666406</td>
<td>4</td>
<td>.10629</td>
</tr>
<tr>
<td>critical</td>
<td>0.0009699</td>
<td>0.0437742</td>
<td>5</td>
<td>.098526</td>
</tr>
<tr>
<td>Always _cons</td>
<td>14.84478</td>
<td>7.871046</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Coefficient posterior means and std. dev. estimated from 32 models.
Note: Default priors are used for models and parameter g.

Although the MC3 sampling is used for the model-space exploration, we can still use `bmagraph coeffdensity` directly after `bmaregress` to plot the analytical posterior densities of regression coefficients. This is because we are still assuming a fixed $g$ parameter in our model.

Let’s compare the analytical and MCMC-sample posterior densities. To obtain the sample estimates, we must first simulate an MCMC sample from the posterior distribution of regression coefficients by using `bmacoefsample`. To use `bmacoefsample`, we first need to save the simulation results produced by `bmaregress`.

```
.bmaregress, saving(bmaex2, replace)
note: file bmaex2.dta not found; file saved.
.bmacoefsample, rseed(18)
Simulation (10000): ....5000.....10000 done
```
To request the density estimate based on the MCMC sample, we specify the `mcmcsample` option. We plot the MCMC-sample posterior density estimate for the `{learning}` coefficient.

```
. bmagraph coefdensity {learning}, mcmcsample
```

The plot is similar to that in example 1.

The marginal density of the `learning` coefficient can also be estimated using the `bayesgraph kdensity` command (see `[BAYES] bayesgraph`). This command, however, incorporates the mass at zero and shows the distribution mixture as one smooth curve, which may not be of practical use in the context of BMA models.
Notice how the density curve shrinks toward 0 given the relatively large probability mass \((1 - \text{PIP} = 0.7477)\) at that point.

For comparison with the density plots in example 1, we will plot MCMC-sample densities for all coefficients in our model. To demonstrate, we also use `dens6opts(lcolor(cyan))` to specify a cyan color for the density on the sixth plot corresponding to `{rating:advance}`.

`bmagraph coefdensity _all, mcmcsample combine(rows(2)) dens6opts(lcolor(cyan))`

If adequate exploration of the model space by `bmaregress` occurred and a sufficiently large MCMC sample is generated by `bmacrofsample`, the analytical and MCMC-sample posterior densities are expected to be similar. This is the case in our examples (after we adjust for the scales of the \(x\) axes).

With a random \(g\), the analytical posterior densities are not available, and the MCMC sampling is needed to approximate these distributions. In that case, the `bmacrofsample` command must be run first to obtain posterior samples of regression coefficients, as we showed above. `bmagraph coefdensity`
can then be used to plot an MCMC-sample posterior density, except we do not need to specify the mcmcsample option—it is implied with a random $g$.

### Methods and formulas

Continuing with the notation in *Methods and formulas* of [BMA] bmaregress, we consider a predictor $X_k$ for $1 \leq k \leq p$. Let $\beta_k$ be the regression coefficient associated with predictor $X_k$.

The posterior distribution of $\beta_k$ has a mixed structure, which consists of a continuous portion (a mixture of location-scale $t$-distributions) and a discrete portion (point mass at 0),

$$p(\beta_k | y) = (1 - \text{PIP}_k) \delta_0 + \text{PIP}_k f_k$$

where $\delta_0$ denotes the point mass at 0 and $f_k$ is defined by (10) in *Posterior distributions of regression coefficients* in *Methods and formulas* of [BMA] bmaregress for models with a fixed $g$ (option analytic). When the mcmcsample option is assumed, either by default for models with a random $g$ or if specified for models with a fixed $g$, $f_k$ is computed using the kdensity (see [R] kdensity) command based on the coefficient sample generated by bmacoefsample.

The vertical line for the probability mass at 0 corresponds to $1 - \text{PIP}_k$, where $\text{PIP}_k$ is the PIP of $X_k$ computed as

$$\text{PIP}_k = \sum_{j \in J} I(X_k \in M_j) P(M_j | y)$$

and $J$ indexes a subspace of the model space visited by bmaregress. Also see *Posterior inclusion probability* in *Methods and formulas* of [BMA] bmaregress.

In the above formula, for models with a fixed $g$ parameter, $P(M_j | y)$ is either $P_a(M_j | y)$ (option analytic) or $P_f(M_j | y)$ (option mcmcsample). For models with a random $g$ parameter, $P(M_j | y)$ is $P_f(M_j | y)$. These quantities are defined, respectively, by (7) and (8) in *Posterior model probability* in *Methods and formulas* of [BMA] bmaregress.

### Reference


### Also see

[BMA] bmagraph — Graphical summary for models and predictors after BMA regression

[BMA] bmaregress — Bayesian model averaging for linear regression

[BMA] BMA postestimation — Postestimation tools for Bayesian model averaging

[BMA] Glossary

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.