A command for estimating regression parameters for the maximum agreement predictor

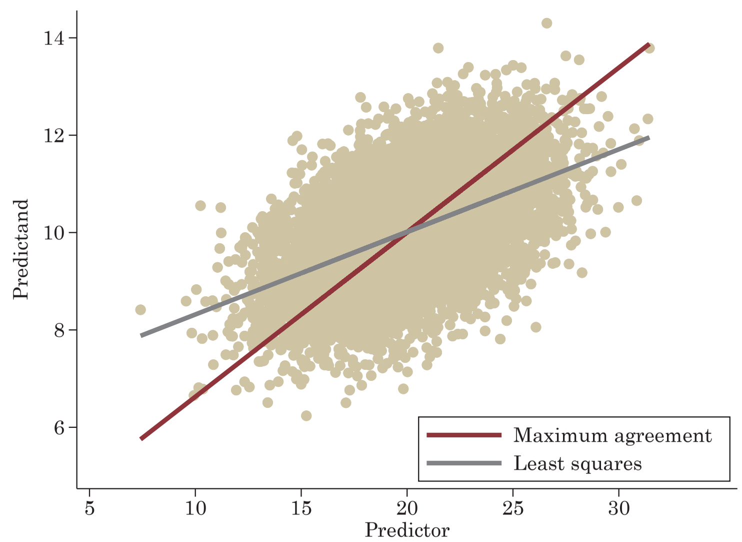
Matteo Bottai, Sc.D.
Karolinska Institutet

Reference

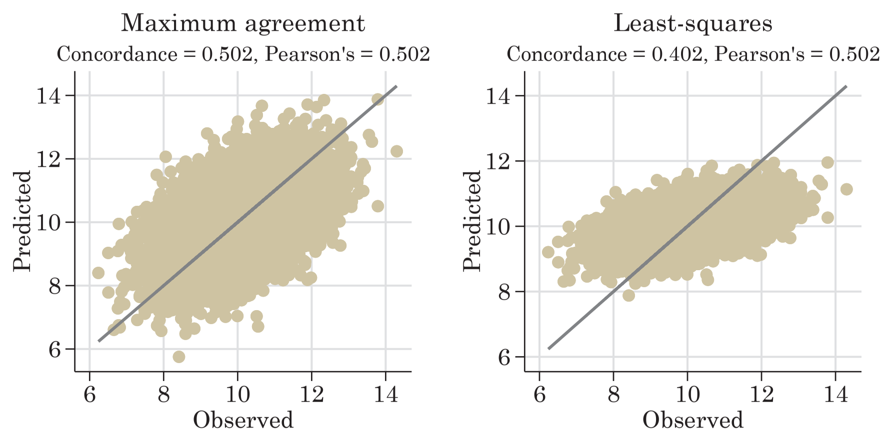
The details of the underlying theory are published in

Bottai M, Kim T, Lieberman B, Luta G, Peña E
On Optimal Correlation-Based Prediction
The American Statistician, 76:4, 313-321, 2022

Maximum agreement and least squares predictors



Predicted values



The do-file

```
. mareg y x
Maximum agreement regression                  Number of obs          =     10000
                                              R-squared              =    0.2518
                                              Concordance corr. coef. =   0.5018

------------------------------------------------------------------------------
             |  Observed   Bootstrap                         Normal-based
           y |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  .3379582    .0027289   123.84   0.000     .3326096    .3433068
       _cons |  3.248312    .0562403    57.76   0.000     3.138083    3.358541
------------------------------------------------------------------------------

. regress y x, noheader
------------------------------------------------------------------------------
           y |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  .1695945    .0029235    58.01   0.000     .1638638    .1753252
       _cons |  6.621815    .0592277   111.80   0.000     6.505717    6.737913
------------------------------------------------------------------------------
```


Syntax

`mareg` [*varlist*] [*if*] [*in*] [*weight*] [, `bsopts`(*string*) *options*]

varlist is the same as that of `regress`.
`bsopts` are the options of the `bootstrap` command.
options are the options of the `regress` command.

Some theory

Bottai et al. considered estimating Y with a predictor $h(X)$.
They assumed $E(Y^2) < \infty$ and defined

$$\mathfrak{H} = \{h : (\mathbb{R}^p, \mathbb{B}^p) \to (\mathbb{R}, \mathbb{B}) \mid E[h(X)^2] < \infty\}$$

with $\mathbb{B}$ and $\mathbb{B}^p$ denoting Borel σ-fields.

Some theory

Bottai et al. defined

$$\tilde{h} = \arg\max_{h \in \mathfrak{H}} \kappa(Y, h(X))$$

with κ denoting a bivariate correlation function.

They showed that

– For Pearson's correlation, the class $\tilde{h}$ is uncountably infinite.
– The MA and LS predictors belong to it.
– The MA predictor is the unique concordance maximizer.

Some theory

Bottai et al. show

$$E(h_{\mathrm{MA}}(X)) = E(Y)$$
$$E(h_{\mathrm{LS}}(X)) = E(Y)$$

$$Var(h_{\mathrm{MA}}(X)) = Var(Y)$$
$$Var(h_{\mathrm{LS}}(X)) \leq Var(Y)$$

$$Cor(Y, h_{\mathrm{MA}}(X)) = Cor^{\mathrm{C}}(Y, h_{\mathrm{MA}}(X))$$
$$Cor(Y, h_{\mathrm{LS}}(X)) \geq Cor^{\mathrm{C}}(Y, h_{\mathrm{LS}}(X))$$

The inequalities are equalities if and only if $Cor(Y, h_{\mathrm{LS}}(X)) = 1$.

Conclusions

The `mareg` command

– estimates maximum agreement regression
– is optimal when the inferential objective is concordance
– shares the syntax of `regress` and `bootstrap`
– is computationally fast
– is available through `ssc` and

```
net from http://www.imm.ki.se/biostatistics/stata
```