Comparing Effects Within and Across Models Using Marginal Effects

The new `mecompare`, `meinequality`, `totalme`, and `suest2` commands

Trenton D. Mize

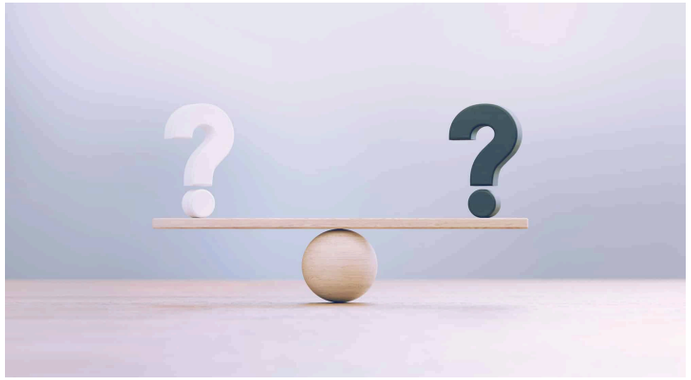
Departments of Sociology & Statistics (by courtesy) · The Methodology Center · Purdue University

2026 Stata Sociology Virtual Symposium · October 20, 2026

PURDUE UNIVERSITY | Methodology Center

Many sociological questions are about comparing effects

1. Within-model comparisons of different predictors
    a. What patterns political views most strongly: age, gender, race-ethnicity, or social class?
2. Tests of interaction
    a. Is the effect of aging on health different for men and women?
3. Mediation/attenuation: cross-model comparisons
    a. How much of education's protective effect on mental health is explained by SES?
4. Tests across samples/groups
    a. Has the role of religion in determining opinions about social issues declined since the 1970s?

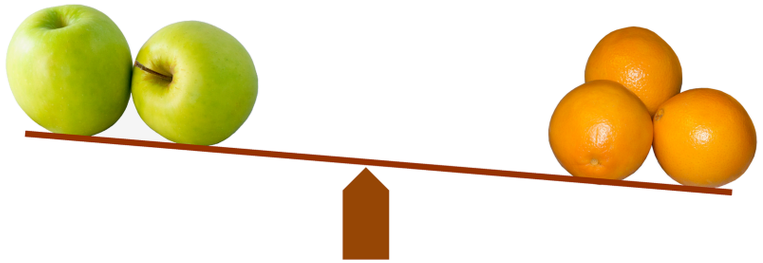


$$\Delta = effect_a - effect_b$$

5. One framework and one set of commands: `mecompare`, `metest`, `suest2`, `meinequality`, and `totalme`

Complications to comparisons and the promise of marginal effects

1. Coefficients are often not in the metric of interest (e.g., `logit`'s log-odds)
2. Predictors are on different metrics: a year of age vs. a dollar of income vs. being a woman
3. Logit and probit coefficients are rescaled when variables are added, even with no confounding
4. Nominal and ordinal variables produce an overwhelming number of comparisons
5. Tests across models need covariances that don't exist when models are fit separately



Overview

1. Marginal effects and `mecompare`
2. Comparing effects within a model
    a. Interactions as comparisons of effects
3. Comparing effects across models: SUEST and `suest2`
4. Summary measures for nominal and ordinal variables: `meinequality` and `totalme`

Marginal effects

What is a marginal effect?

1. Your model's answer to a what-if question: if only x_k changed, how much would the prediction (η) change, with the other variables ($\mathbf{x}_{-k}$) held constant?

$$ME = \eta\left(x_k = end,\ \mathbf{x}_{-k} = \mathbf{x}^*\right) - \eta\left(x_k = start,\ \mathbf{x}_{-k} = \mathbf{x}^*\right)$$

2. You choose the change (a slope, +1, +SD, a contrast between categories) and the prediction (probabilities, counts, means, ...)
3. You choose where the other variables stay: observed values (then averaged: the AME, the default in `mecompare`), means, or values of interest

$$AME = \frac{1}{N}\sum_{i=1}^{N}\left\{\eta\left(x_k = end_i,\ \mathbf{x}_{-k} = \mathbf{x}_{-k,i}\right) - \eta\left(x_k = start_i,\ \mathbf{x}_{-k} = \mathbf{x}_{-k,i}\right)\right\}$$

4. "Effect" means a change implied by the model; whether it is causal depends on the design

Testing the difference between two marginal effects

$$z = \frac{\widehat{ME}_a - \widehat{ME}_b}{\sqrt{\widehat{\sigma}_a^2 + \widehat{\sigma}_b^2 - 2\,\widehat{\sigma}_{a,b}}}$$

1. a and b can be:
    a. Two predictors in one model
    b. One predictor at two values of a moderator (interaction)
    c. One predictor in two models
2. Requires the covariance $\widehat{\sigma}_{a,b}$
    a. Within one model: from the model's covariance matrix
    b. Across models: not available when the models are fit separately

mecompare

```
mecompare [varlist] [if] [in] [weight] [, models(name1 [name2 ...]) options]
```

1. Marginal effects for one model or many
    a. With two or more, `suest2` combines the models' estimates
2. Binary and nominal contrasts (to a reference, all pairwise, summaries)
3. Continuous predictors with any amount of change (`dydx`, +1, +SD, min to max, …)
4. `metest` tests any comparison of the effects in the table
5. Options for interactions (`by()`, `over()`, `covariates()`), groups, and summary measures

Comparing effects within a model

Binary vs. continuous predictors

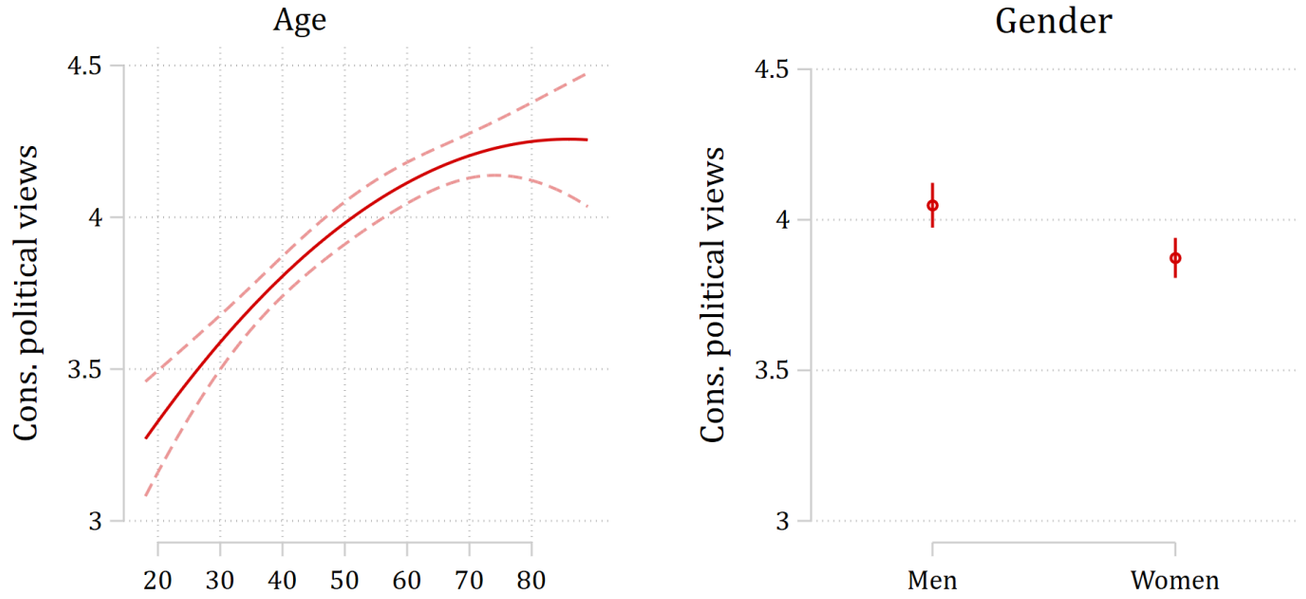
1. A binary predictor's 0 to 1 change is its whole range; what is comparable for a continuous predictor?
    a. +1 is arbitrary and rarely a meaningful amount
2. +2SD (Gelman 2008): centered on each observed value; comparable to a 50/50 binary variable

$$ME_i^{+2SD} = \eta\left(x_k = x_{k,i} + SD,\ \mathbf{x}_{-k,i}\right) - \eta\left(x_k = x_{k,i} - SD,\ \mathbf{x}_{-k,i}\right)$$

3. Trimmed range (Mize & Han 2025): the 5th to the 95th percentile; the whole range, but robust to outliers

$$ME_i^{trim} = \eta\left(x_k = x_k^{p95},\ \mathbf{x}_{-k,i}\right) - \eta\left(x_k = x_k^{p5},\ \mathbf{x}_{-k,i}\right)$$

Example: what patterns political views?



1. GSS 2021: political views (1 to 7, conservative high), linear regression
2. Older adults are more conservative; nonlinear effect (age^2 $p < .05$)
3. Women are slightly less conservative than men
4. Which matters more?

The effects on one metric

1. `amount()` sets the change: `one` (default), `sd`, `2sd`, `trimrange`, `range`, `p10-p90`, `rate`, or any number
2. Factor syntax is required on the stored model (`i.woman`)

```
quietly regress conpolviews c.age##c.age i.woman i.race4 i.class, vce(robust)
estimates store polmod
mecompare age woman, models(polmod) amount(trimrange)
```

```
Predicting: Linear prediction
Marginal effects (N_polmod=3608)

                                 |  ME #   Estimate  Robust SE      P>|z|
---------------------------------+-----------------------------------------
age(5-95%)                       |
                          polmod |     1      0.784      0.083      0.000
---------------------------------+-----------------------------------------
woman                            |
                  Women - Men    |
                          polmod |     2     -0.175      0.051      0.001
```

Across its trimmed range (age 25 to 79), age increases conservatism by 0.78 points. Women are 0.18 points less conservative than men.

Testing the difference in effect size

1. Is the effect of age larger than the effect of gender?
2. In `metest`, each number is a row (ME #) of the `mecompare` table
3. `abs()` compares sizes regardless of sign

```
metest 1 - abs(2)
```

```
                                |  estimate          se     pvalue
--------------------------------+-----------------------------------
               age - abs(woman) |     0.609       0.098      0.000
```

The age effect is larger than the gender gap by 0.61 points ($p < 0.001$).

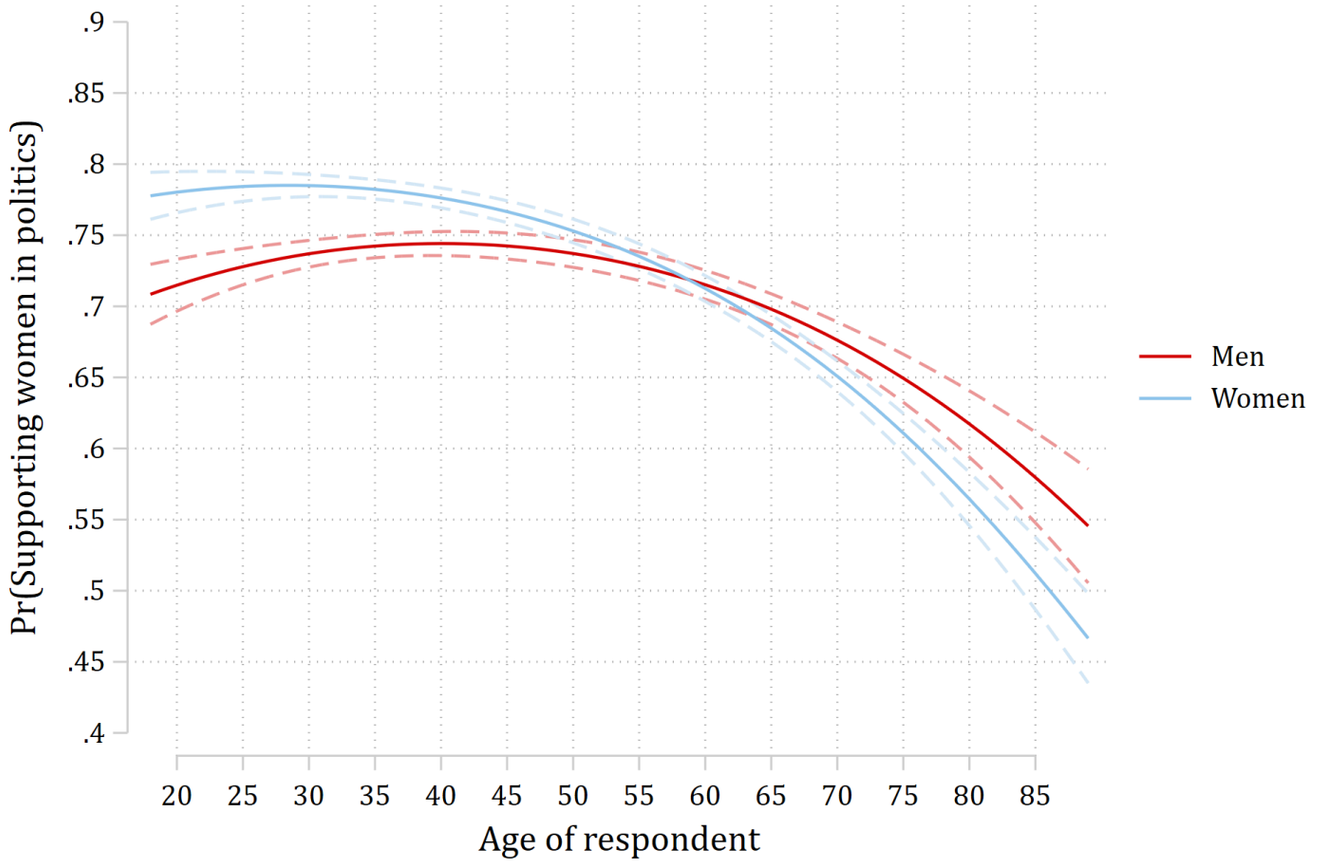
Interactions are comparisons of effects

Interactions as comparisons of effects

$$\Delta = ME_{x_k|z=1} - ME_{x_k|z=0}$$

1. A test of interaction asks: is the effect of x_k different at different values of z?
    a. A comparison of two MEs from one model: the same Wald test
2. In nonlinear models, the product term's coefficient is not that test: include the term, but test the MEs (Mize 2019)
3. Plot the predictions, compute MEs at levels of the moderator (first differences), then test their equality (second difference)
4. Nominal moderator: test at all values
    a. Continuous moderator: MEs at meaningful values, or the 10th, 50th, and 90th percentiles
5. `by()` in `mecompare` automates tests of interaction (`over()` uses group-specific observations)

Example: support for women in politics



1. GSS: men and women equally suited for political careers
    a. Logit with age, age^2, gender, and interactions
2. Young women are more supportive than young men
3. The gap closes around age 60 and reverses after
4. Is the effect of aging different for men and women?

Marginal effects of age for men and women

1. +5 years of age (`amount(5) uncentered`) from ages 25 and 72, the 10th and 90th percentiles (`start()`), for men and women (`by(woman)`)

```
quietly logit fepolB c.age##c.age##i.woman, vce(robust)
estimates store agemod
quietly summarize age, detail
mecompare age, models(agemod) start(age=(`r(p10)' `r(p90)')) amount(5) uncentered by(woman)
```

```
Predicting: Pr(fepolB)
Marginal effects (N_agemod=36712)

                                |  ME #   Estimate  Robust SE      P>|z|
--------------------------------+-----------------------------------------
age + 5 (uncentered)            |
                     Men at 25  |     1      0.009      0.003      0.001
                     Men at 72  |     2     -0.029      0.004      0.000
                   Women at 25  |     3      0.001      0.002      0.754
                   Women at 72  |     4     -0.043      0.003      0.000
```

At 25, aging five years increases support among men (0.009) but not women (0.001); at 72, it lowers support for both, potentially more for women (−0.043) than men (−0.029).

Second differences

1. `metest ..., add` builds a table of tests; `rowname()` labels each row; `title()` shows the table; `metest, clear` starts a new one

```
metest, clear
metest 3 - 1, add rowname("2nd diff at age 25")
metest 4 - 2, add rowname("2nd diff at age 72")
metest, title("Second differences: women - men")
```

```
Second differences: women - men

                                |  estimate         se     pvalue
--------------------------------+-----------------------------------
             2nd diff at age 25 |    -0.008      0.003      0.015
             2nd diff at age 72 |    -0.014      0.005      0.004
```

At both ages, the effect of aging is smaller for women than for men (−0.008 at 25 and −0.014 at 72), indicating an age by gender interaction.

The other side: the gender gap at three ages

1. A list in `covariates()`: one ME per value, each as if the whole sample were that age
2. `metest 1 = 2 = 3`: a joint test that the gap is the same at all three ages (passed to `test`)

```
quietly summarize age, detail
mecompare woman, models(agemod) covariates(age=(`r(p10)' `r(p50)' `r(p90)'))
metest 1 = 2 = 3
```

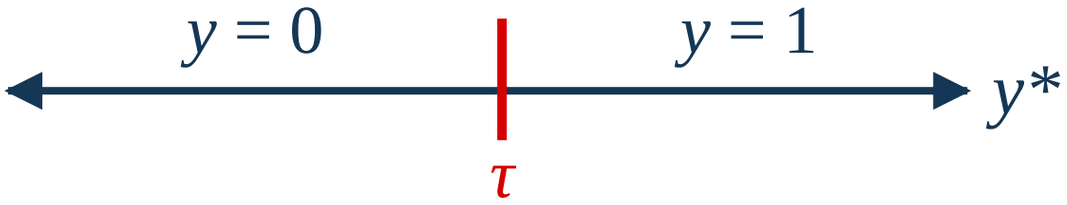
```
                                 |  ME #   Estimate  Robust SE      P>|z|
---------------------------------+-----------------------------------------
woman                            |
                  Women - Men    |
                           age=25 |     1      0.056      0.008      0.000
                           age=44 |     2      0.026      0.006      0.000
                           age=72 |     3     -0.031      0.009      0.001
```

```
                                 |      chi2         df     pvalue
---------------------------------+-------------------------------
  woman~25 = woman~44 = woman~72 |    45.251      2.000      0.000
```

The gender gap varies across young, middle, and older ages (χ^2 = 45.25, df = 2, p < 0.001), showing an interaction between age and gender.

Comparing effects across models

Rescaling in nonlinear models





$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i, \qquad y_i = 1 \text{ if } y_i^* > \tau$$

$$\beta_k^{\text{logit}} = \beta_k \frac{\pi/\sqrt{3}}{\sigma_\varepsilon}$$

1. Logit and probit fix the error variance ($\pi^2/3$ for logit), which sets the scale of y^*
2. A predictor that explains y^* shrinks σ_ε, so coefficients grow, even if it is unrelated to x_k
3. Coefficient changes across models mix confounding and rescaling (Karlson, Holm & Breen 2012)
4. MEs on the probability are not rescaled, so they can be compared across models (Mize, Doan & Long 2019)

Cross-model comparisons

$$\Delta_m = ME_k^{(m_1)} - ME_k^{(m_2)} \qquad \text{Proportion change} = \frac{ME_k^{(m_1)} - ME_k^{(m_2)}}{ME_k^{(m_1)}}$$

1. The same Wald test, but the covariance is across models

$$z = \frac{\widehat{ME}_k^{(m_1)} - \widehat{ME}_k^{(m_2)}}{\sqrt{\widehat{\sigma}^2_{m_1} + \widehat{\sigma}^2_{m_2} - 2\,\widehat{\sigma}_{m_1,m_2}}}$$

2. Fit separately, the models give no estimate of that covariance
    a. Assuming 0 gives incorrect tests when the models share observations
    b. Even separate samples can covary (clustering, stratification)
3. Applications: mediation/attenuation, alternative specifications, outcomes, model types, groups (Mize, Doan & Long 2019)

Seemingly unrelated estimation (SUEST)

1. Combines separately fit models into one set of results (Weesie 1999)
    a. Same coefficients, plus the cross-model covariances
2. Intuition: stack the data, one copy per model, fit the models jointly, and cluster on person
3. In practice: combines the models' equation-level scores; robust (sandwich) standard errors
    a. Cross-model covariance for models p and q

$$\widehat{\mathrm{Cov}}\left(\hat{\boldsymbol{\beta}}_p, \hat{\boldsymbol{\beta}}_q\right) = \mathbf{D}_p^{-1}\left(\sum_i w_i\, \mathbf{u}_{pi}\, \mathbf{u}_{qi}'\right)\mathbf{D}_q^{-1}$$



Extending SUEST to multilevel and longitudinal models

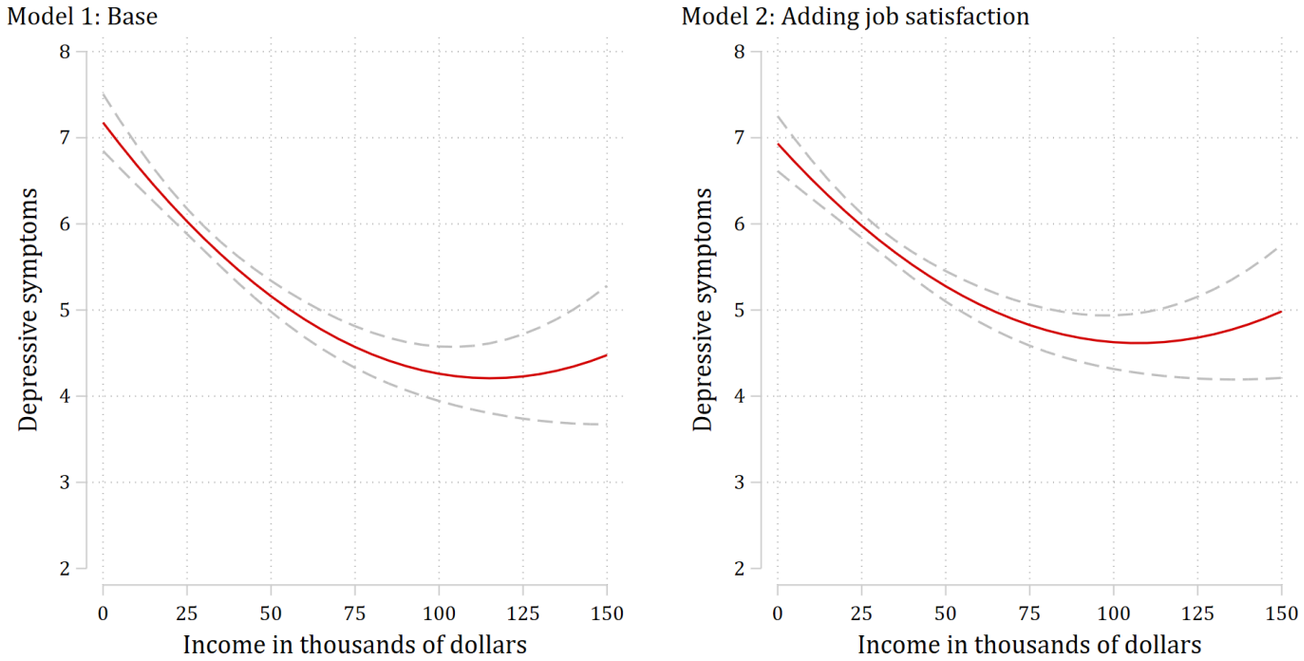
1. `suest2` uses `suest`'s calculation but posts results that work with `margins`
2. `suest` does not support panel (`xt`) or multilevel (`me`) models, whose likelihoods are defined for groups[1]
3. `suest2` builds each model's scores ($\mathbf{u}_{pg}$) for each highest-level group g (persons, schools)

$$\widehat{\mathrm{Cov}}\left(\hat{\boldsymbol{\beta}}_p, \hat{\boldsymbol{\beta}}_q\right) = \mathbf{D}_p^{-1}\left(\sum_{g=1}^{G} \mathbf{u}_{pg}\,\mathbf{u}_{qg}'\right)\mathbf{D}_q^{-1}$$

4. New models in `suest2` were verified against `gsem`, stacked data, and the bootstrap

1. For recommended model specification for fixed effects estimates, see the Appendix

Example: a curvilinear effect and attenuation



1. Add Health Wave 4: depressive symptoms on income and income^2
2. Does job satisfaction account for part of the income effect?
    a. Two coefficients, one summary AME

Reading the table

1. With two models, `mecompare` combines them with `suest2` and reports each model's ME, their difference, and its test

```
quietly regress depsympB c.income##c.income c.age i.woman i.race, vce(robust)
estimates store basemod
quietly regress depsympB c.income##c.income c.age i.woman i.race i.jobsat, vce(robust)
estimates store medmod
mecompare income, models(basemod medmod) amount(sd)
```

```
Predicting: Linear prediction
Marginal effects and cross-model differences (N_basemod=4307) (N_medmod=4307)

                                 |  ME #   Estimate  Robust SE      P>|z|
---------------------------------+-------------------------------------------
income + SD (centered)           |
                         basemod |     1     -0.982      0.098      0.000
                          medmod |     2     -0.797      0.096      0.000
                      Difference |     3     -0.184      0.027      0.000
```

ME #1, base model: a SD increase in income lowers depressive symptoms by 0.98, on average
ME #2, adding job satisfaction: by only 0.80
ME #3, the difference: 0.18 smaller with job satisfaction in the model ($p < 0.001$)

Custom tests and combinations of MEs with `metest`

1. `metest` passes calculations to `test` or `nlcom`, so nonlinear combinations are allowed

```
Predicting: Linear prediction
Marginal effects and cross-model differences (N_basemod=4307) (N_medmod=4307)

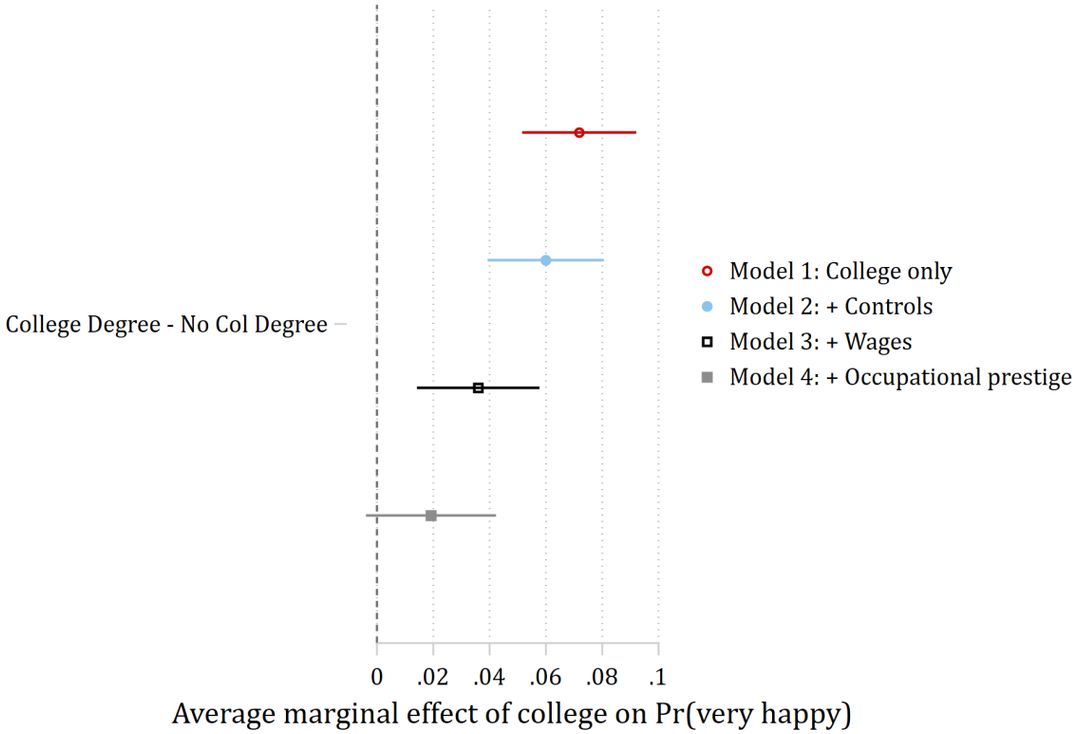
                                  |  ME #   Estimate  Robust SE      P>|z|
----------------------------------+-----------------------------------------
income + SD (centered)            |
                         basemod  |     1     -0.982      0.098      0.000
                          medmod  |     2     -0.797      0.096      0.000
                      Difference  |     3     -0.184      0.027      0.000
```

```
metest (1 - 2) / 1
```

```
                                  |  estimate         se     pvalue
----------------------------------+-----------------------------------
income                            |
    (basemod - medmod) / basemod  |     0.188      0.030      0.000
```

The proportion change: job satisfaction accounts for about 19% of the association between income and depressive symptoms.

Example: four nested logit models



1. GSS 2000 to 2016, employed adults (Mize, Doan & Long 2019)
2. Four logits of being very happy, adding controls, wages, and occupational prestige
   a. `store()` saves the MEs for `coefplot`, `esttab`, etc.
3. The college effect shrinks as each set of variables is added

```
mecompare college, models(m1 m2 m3 m4) store(nest)
coefplot nest_m1 nest_m2 nest_m3 nest_m4, xline(0)
```

Testing across models

1. With three or more models, `mecompare` lists each model's MEs; `metest` tests any pair

```
mecompare college, models(m1 m2 m3 m4) store(nest)
```

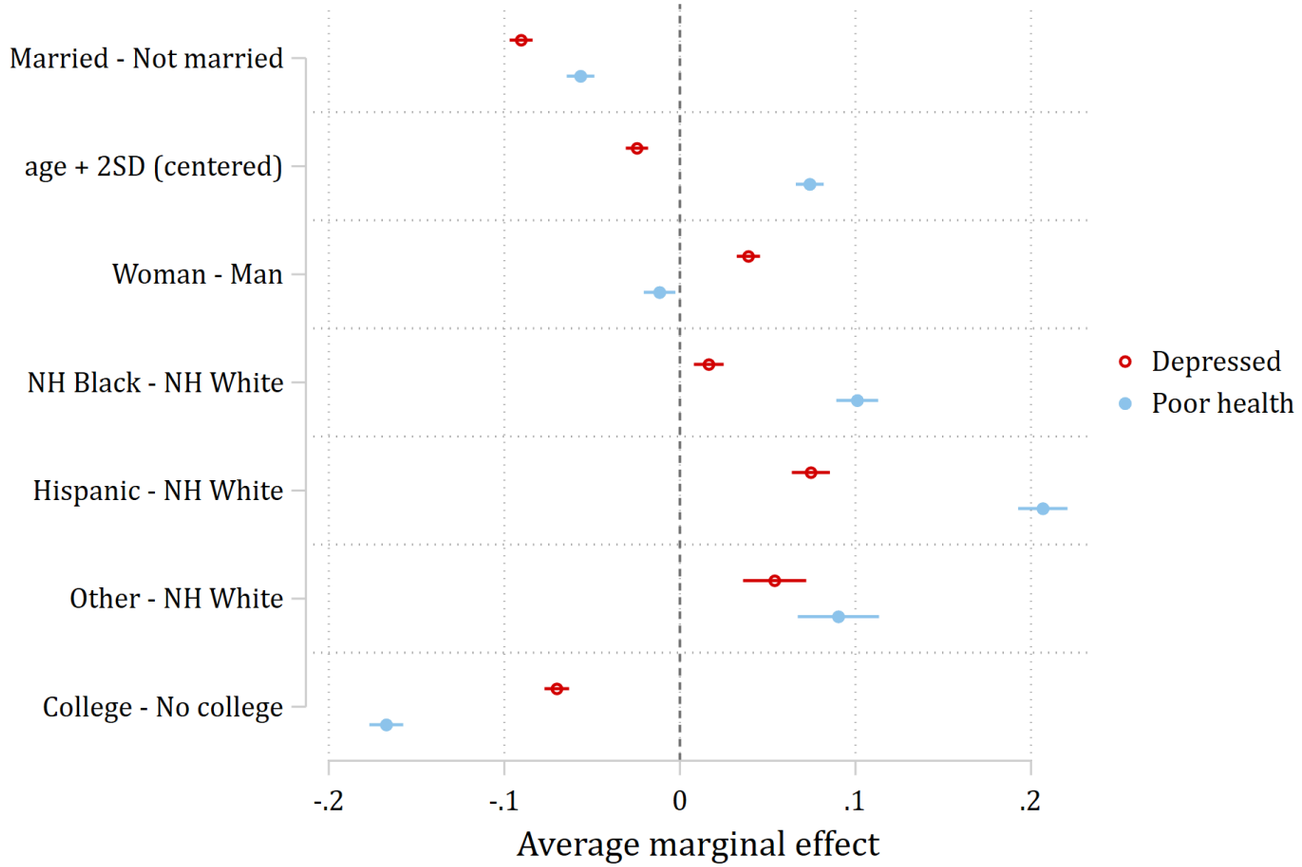
```
                                |  ME #   Estimate  Robust SE      P>|z|
--------------------------------+-------------------------------------
college                         |
    College Deg - No Col Deg    |
                           m1   |     1      0.072      0.010      0.000
                           m2   |     2      0.060      0.011      0.000
                           m3   |     3      0.036      0.011      0.001
                           m4   |     4      0.019      0.012      0.103
```

```
metest, clear
metest 1 - 2, add
metest 2 - 3, add
metest 3 - 4, add
```

```
                                |  estimate         se     pvalue
--------------------------------+--------------------------------
college                         |
                       m1 - m2  |     0.012      0.004      0.003
                       m2 - m3  |     0.024      0.004      0.000
                       m3 - m4  |     0.017      0.004      0.000
```

The college effect falls from 0.072 to 0.019; each set of added variables reduces it significantly; the ME is no longer significant in the final model.

Example: longitudinal models



1. HRS 2010 to 2020, adults 51 and older
2. Random-effects logits (`xtlogit, re`) of depression and of poor health
3. Most predictors matter more for poor health than for depression
4. Exceptions: marriage and gender matter more for depression

Longitudinal models: tests

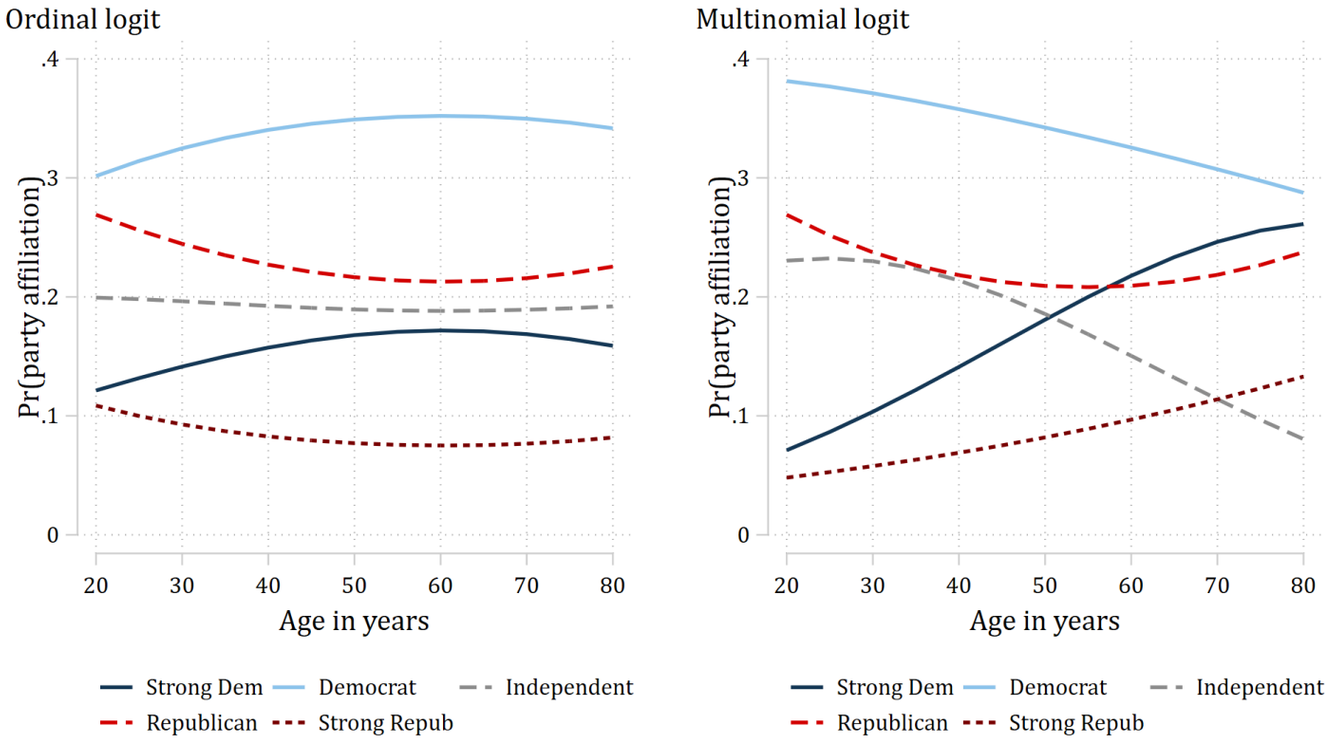
```
quietly xtlogit depressed i.marstatB c.age i.woman i.race4cat i.collegeB, re intpoints(60)
estimates store dep
quietly xtlogit poorhealth i.marstatB c.age i.woman i.race4cat i.collegeB, re intpoints(60)
estimates store health
mecompare marstatB collegeB, models(dep health)
```

```
                                |  ME #   Estimate  Robust SE      P>|z|
--------------------------------+-------------------------------------
marstatB                        |
       Married - Not married    |
                           dep  |     1     -0.090      0.003      0.000
                        health  |     2     -0.057      0.004      0.000
                    Difference  |     3     -0.034      0.004      0.000
--------------------------------+-------------------------------------
collegeB                        |
        College - No college    |
                           dep  |     4     -0.070      0.004      0.000
                        health  |     5     -0.167      0.005      0.000
                    Difference  |     6      0.097      0.005      0.000
```

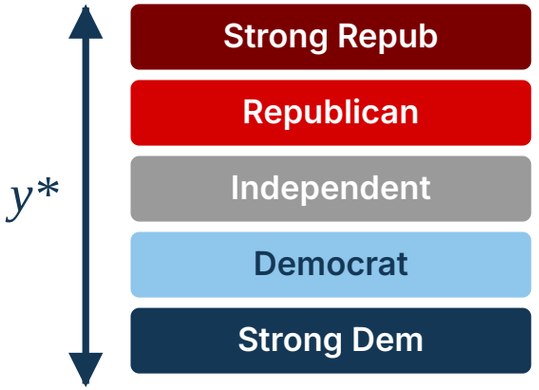
No `vce(robust)` on the models: `mecompare` clusters the standard errors on person.

Being married matters more for depression (−0.090) than for poor health (−0.057); a college degree matters more for poor health (−0.167 vs. −0.070).

Example: different model types



1. GSS 2010 to 2016: party identification (5 categories)
2. The same model as an ordinal and as a multinomial logit
3. The ordinal model forces age to shift all categories in one direction
    a. Does that change the conclusions?



Different model types: tests

1. Any models with the same number of outcome categories can be compared (here, `ologit` vs. `mlogit`)

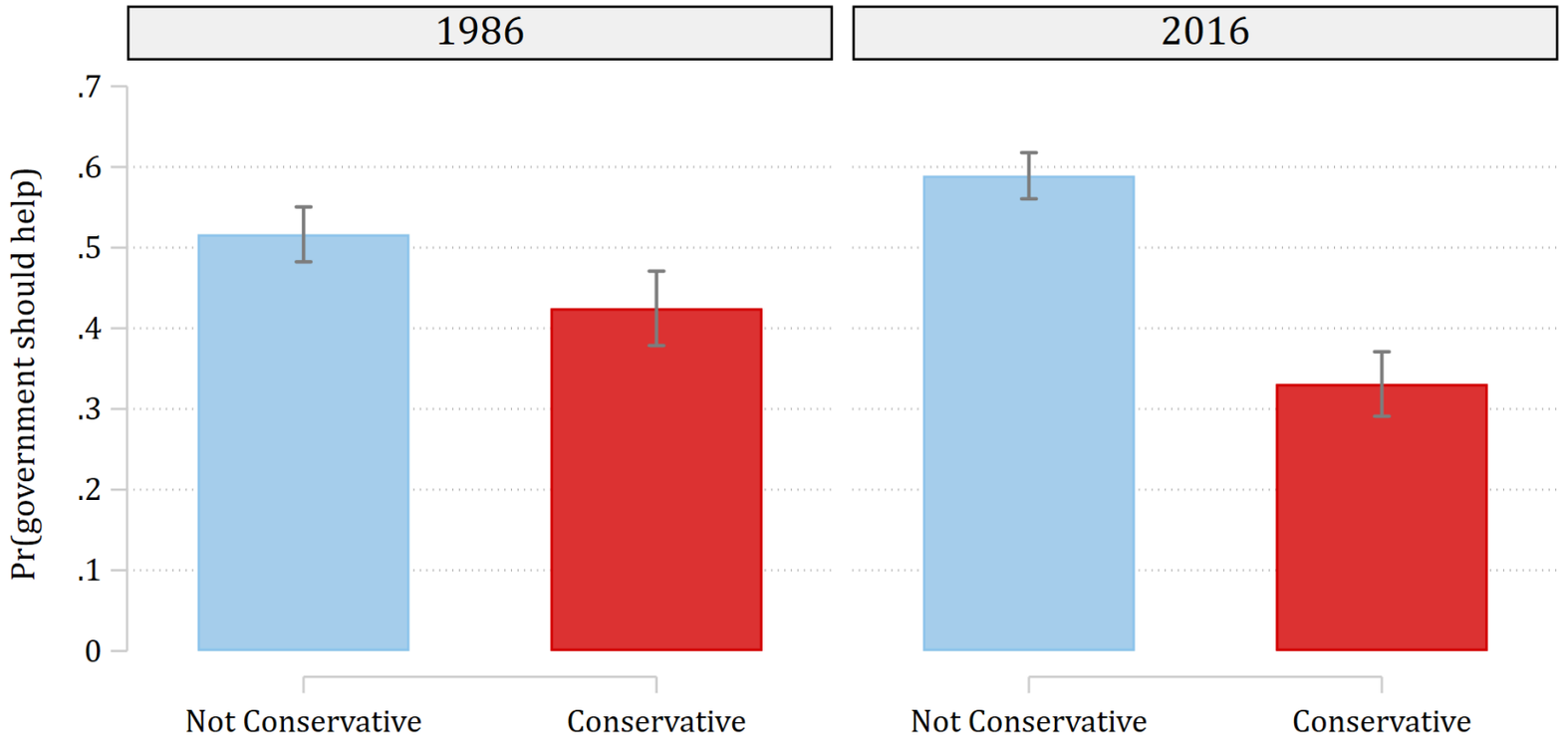
```
mecompare age, models(ordinal nominal) amount(sd)
```

```
Predicting: Pr(partyid5)
Marginal effects and cross-model differences (N_ordinal=8179) (N_nominal=8179)

                                 |  ME #   Estimate  Robust SE      P>|z|
---------------------------------+-----------------------------------------
age + SD (centered)              |
            Strong Dem - ordinal |     1      0.014      0.003      0.000
            Strong Dem - nominal |     2      0.054      0.004      0.000
         Strong Dem - Difference |     3     -0.040      0.003      0.000
              Democrat - ordinal |     4      0.009      0.002      0.000
              Democrat - nominal |     5     -0.029      0.006      0.000
           Democrat - Difference |     6      0.038      0.006      0.000
           Independent - ordinal |     7     -0.002      0.000      0.000
           Independent - nominal |     8     -0.040      0.005      0.000
        Independent - Difference |     9      0.038      0.005      0.000
            Republican - ordinal |    10     -0.012      0.003      0.000
            Republican - nominal |    11     -0.011      0.006      0.050
         Republican - Difference |    12     -0.001      0.005      0.828
          Strong Repub - ordinal |    13     -0.009      0.002      0.000
          Strong Repub - nominal |    14      0.027      0.004      0.000
       Strong Repub - Difference |    15     -0.036      0.003      0.000
```

The models disagree for 4 of 5 outcomes: the multinomial model shows aging linked mainly to strong affiliations on either side; the ordinal model cannot show this, so its results are biased.

Example: groups and separate samples



1. GSS, separate logits for 1986 and 2016: should the government help pay for medical care?
2. Conservatives are less supportive in both years, and the gap appears wider in 2016

Groups and separate samples: tests

1. `groups`: each model is fit to its own sample (here, `if year == ...`); the samples must not overlap

```
quietly logit helpsickB i.conserv faminc i.employed i.woman age i.college ///
    i.married i.parent i.race if year == 1986, vce(robust)
estimates store mod1986
quietly logit helpsickB i.conserv faminc i.employed i.woman age i.college ///
    i.married i.parent i.race if year == 2016, vce(robust)
estimates store mod2016
mecompare conserv, models(mod1986 mod2016) groups
```

```
                                |  ME #   Estimate  Robust SE      P>|z|
--------------------------------+-----------------------------------------
conserv                         |
    Conservativ - Not Conser    |
                       mod1986  |     1     -0.092      0.030      0.002
                       mod2016  |     2     -0.258      0.025      0.000
                     Difference |     3      0.166      0.039      0.000
```

Conservatives' lower support grew from 0.092 in 1986 to 0.258 in 2016; the 0.166 difference suggests growing polarization on this issue ($p < 0.001$).

Summary measures for nominal and ordinal variables

Many effects, one question

1. A nominal predictor with J categories has $J(J-1)/2$ pairwise contrasts (religion, 7 categories: 21)
2. Reference-category coefficients test only $J-1$ of them and privilege one group (Johfre & Freese 2021)
3. Holistic questions: how much does race-ethnicity matter? Has religion's effect declined?
4. Summary measures: ME inequality and the Total ME (Mize & Han 2025)

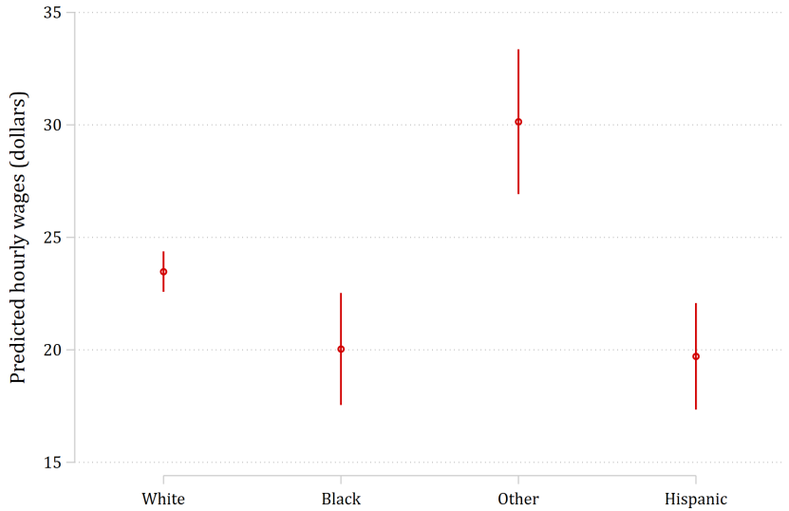


ME inequality

$$\text{ME inequality} = \sum_{a<b} \frac{p_a + p_b}{J-1} \left| \eta\left(x_k = a\right) - \eta\left(x_k = b\right) \right|$$

1. The average absolute difference in predictions across the categories of x_k: no reference category
2. Each pair weighted by its share of the sample, $p_a + p_b$; with `unweighted`, each pair counts equally
3. Delta-method standard errors via `nlcom`; compare across models and groups with `suest2`
4. The `meinequality` command, or `mecompare ..., meinequality`

Example: racial-ethnic inequality in wages



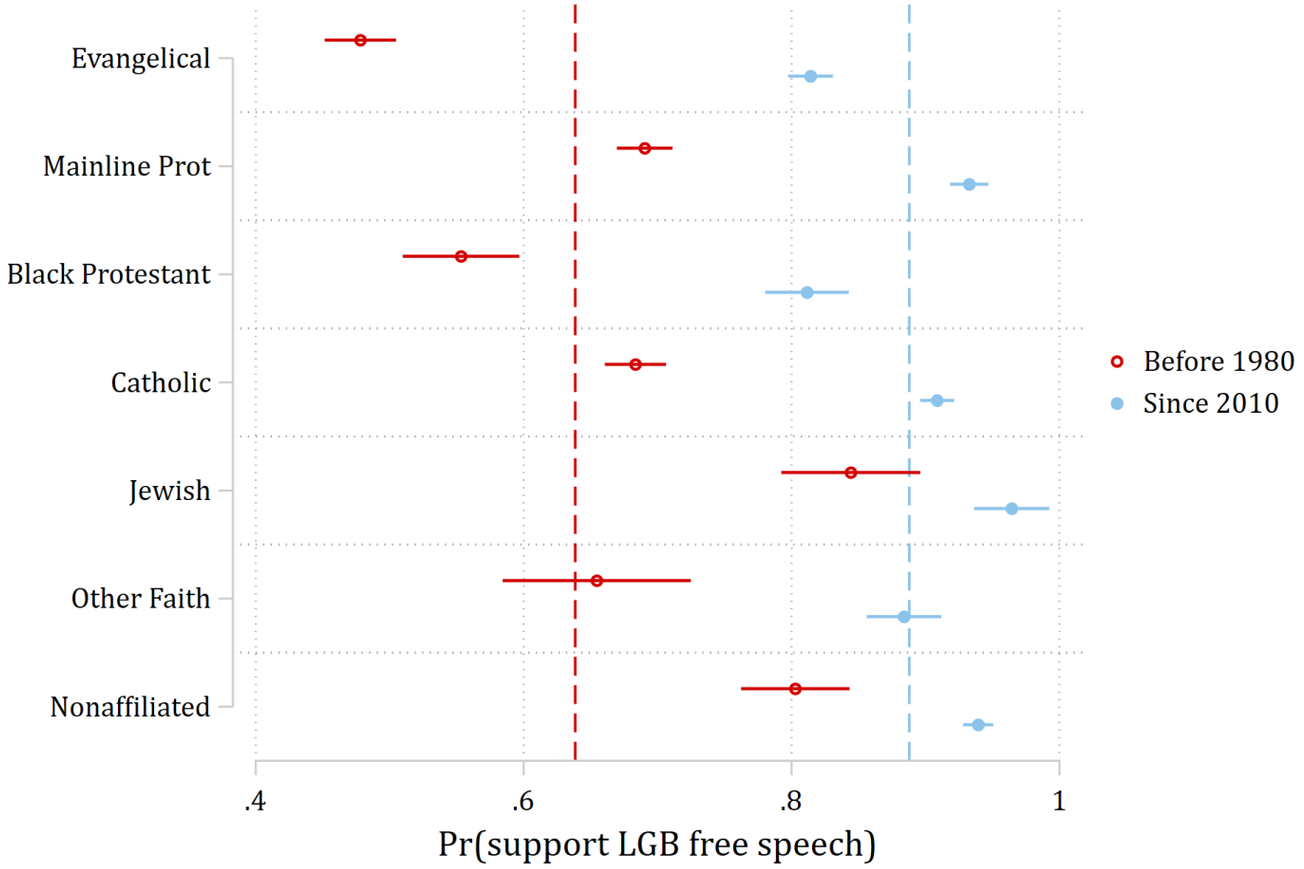
1. GSS 2021: hourly wages by race-ethnicity, net of age and gender
2. Predicted wages range from $19.71 (Hispanic) to $30.14 (Other)
3. How much does race-ethnicity matter overall?

```
quietly regress wages i.race4 c.age i.woman
meinequality race4
```

```
                         |   Estimate   Std. err.           z       P>|z|
-------------------------+----------------------------------------------
ME Inequality            |
                   race4 |      4.904       0.860       5.702       0.000
```

On average, racial-ethnic groups' predicted wages differ by about $4.90 per hour.

Example: has religion's effect declined?



1. GSS: allowing a gay person to give a public speech, by religion (7 traditions)
2. Separate logits before 1980 and since 2010
3. Support rose in every tradition, and the gaps narrowed
4. How much did religion's holistic effect decline?

`meinequality` across samples

1. `models()` and `groups` work as in `mecompare`; the models are combined with `suest2`

```
quietly logit spkhomo i.reltrad c.age i.woman if year < 1980, vce(robust)
estimates store premod
quietly logit spkhomo i.reltrad c.age i.woman if year >= 2010, vce(robust)
estimates store postmod
meinequality reltrad, models(premod postmod) groups unweighted
```

```
ME Inequality Estimates (N_premod = 5593 , N_postmod = 8226)

                         |   Estimate   Std. err.           z       P>|z|
-------------------------+----------------------------------------------
reltrad Unwgt ME Ineq.   |
        Model 1 (premod) |      0.156       0.011      14.329       0.000
       Model 2 (postmod) |      0.072       0.007      10.913       0.000
       Cross-Model Diff. |      0.083       0.013       6.561       0.000
```

Religious traditions differed by an average probability of 0.156 before 1980 and 0.072 since 2010: religion's effect declined by about half.

Total marginal effect

1. With L outcome categories (`mlogit`, `ologit`, `oprobit`, ...), each predictor has L MEs: one per category
2. The Total ME sums them into one number: the total probability a predictor moves across the categories

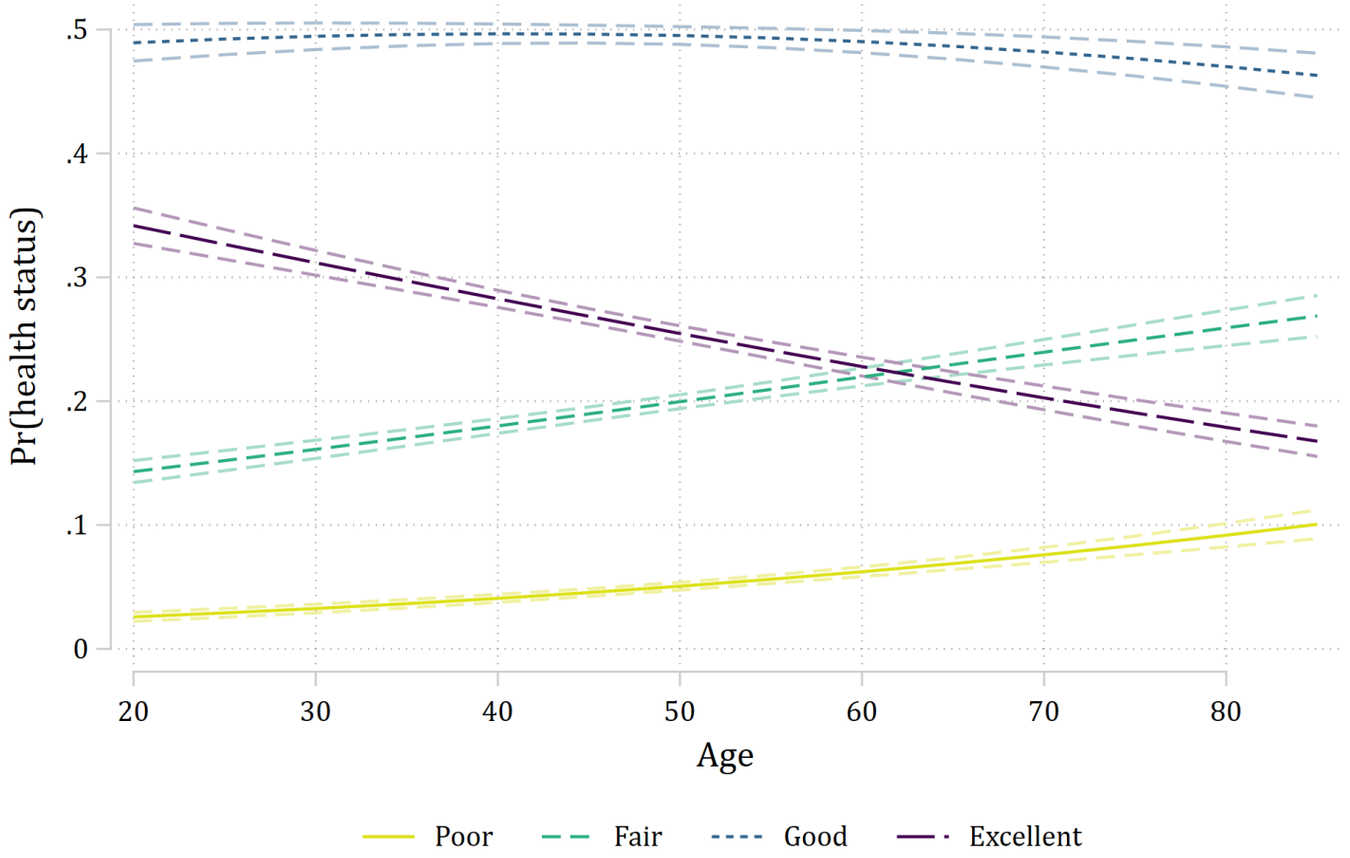
$$\text{Total ME}_k = \frac{1}{2} \sum_{l=1}^{L} |ME_{k,l}|$$

3. Halved because MEs are zero-sum
4. Nominal predictor: the Total ME inequality, with the ME inequality weights

$$\text{Total ME inequality} = \sum_{a<b} \frac{p_a + p_b}{J - 1} \; \frac{1}{2} \sum_{l=1}^{L} |ME_{ab,l}|$$

5. The `totalme` command, or `mecompare ..., totalme`

Example: self-rated health



1. GSS 2000 to 2021: self-rated health (4 categories), multinomial logit
2. With age, poor and fair health rise and excellent health falls
3. How much does age matter overall, vs. marriage and parenthood?

totalme

```
quietly mlogit healthR i.race4 c.age i.woman i.parent i.married c.faminc i.degree
totalme age married parent, amount(sd)
```

```
Total ME Estimates (N = 19292)

                         |   Estimate   Std. err.           z       P>|z|
-------------------------+----------------------------------------------
age                      |
         + SD (centered) |      0.051       0.003      16.847       0.000
-------------------------+----------------------------------------------
married                  |
  Married vs Not Married |      0.031       0.007       4.737       0.000
-------------------------+----------------------------------------------
parent                   |
   Parent vs No Children |      0.027       0.009       2.932       0.003
```

A SD increase in age moves about 5% of the sample into a different health category; marriage and parenthood each move about 3%.

1. To test whether these total MEs differ, use `mecompare` and `metest`

Thank you

Trenton D. Mize · tmize@purdue.edu

Slides, documentation, and worked examples: trentonmize.com/software

```
net install suest2, from("https://tdmize.github.io/data") replace
net install mecompare, from("https://tdmize.github.io/data") replace
net install meinequality, from("https://tdmize.github.io/data") replace
net install totalme, from("https://tdmize.github.io/data") replace
```

Requires Stata 16 or later; `metest` installs with `mecompare`.

> **Many sociological questions involve comparing effects. Marginal effects provide a consistent framework to test effects within or across almost any model.**

Questions?

References

Gelman, Andrew. 2008. “Scaling Regression Inputs by Dividing by Two Standard Deviations.” *Statistics in Medicine* 27(15): 2865–2873.

Johfre, Sasha Shen, and Jeremy Freese. 2021. “Reconsidering the Reference Category.” *Sociological Methodology* 51(2): 253–269.

Karlson, Kristian Bernt, Anders Holm, and Richard Breen. 2012. “Comparing Regression Coefficients Between Same-Sample Nested Models Using Logit and Probit: A New Method.” *Sociological Methodology* 42(1): 286–313.

Mize, Trenton D. 2019. “Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects.” *Sociological Science* 6: 81–117.

Mize, Trenton D., Long Doan, and J. Scott Long. 2019. “A General Framework for Comparing Predictions and Marginal Effects Across Models.” *Sociological Methodology* 49(1): 152–189.

Mize, Trenton D. and Bing Han. 2025. “Inequality and Total Effect Summary Measures for Nominal and Ordinal Variables.” *Sociological Science* 12: 115–157.

Mize, Trenton D. and Bing Han. Forthcoming. Marginal effects: flexible methods for interpretation across linear and nonlinear models. In *Understanding Data Modeling and Data Analysis*, edited by David Weakliem. Edward Elgar Publishing.

Weesie, Jeroen. 1999. “Seemingly Unrelated Estimation and the Cluster-Adjusted Sandwich Estimator.” *Stata Technical Bulletin* 52: 34–47.

Appendix

Generalized marginal effects

1. Co-change variables move along with the focal variable instead of being held constant
2. Here: wages and occupational prestige change from their means for those without to those with a college degree

```
mecompare college, models(m4) cochange(wages occprest) ///
    start(wages=13.60 occprest=40.51) end(wages=24.88 occprest=54.68)
```

```
                                |  ME #   Estimate  Robust SE      P>|z|
--------------------------------+-------------------------------------
college                         |
    College Deg - No Col Deg    |
                           m4   |     1      0.019      0.012      0.103
--------------------------------+-------------------------------------
college with co-change          |
                           m4   |     2      0.063      0.011      0.000

NOTE: college with co-change: college 0 to 1; wages 13.60 to 24.88; occprest 40.51 to 54.68.
```

Holding wages and prestige constant, the college effect is 0.019; letting them change with college, it is 0.063.

Within-person effects

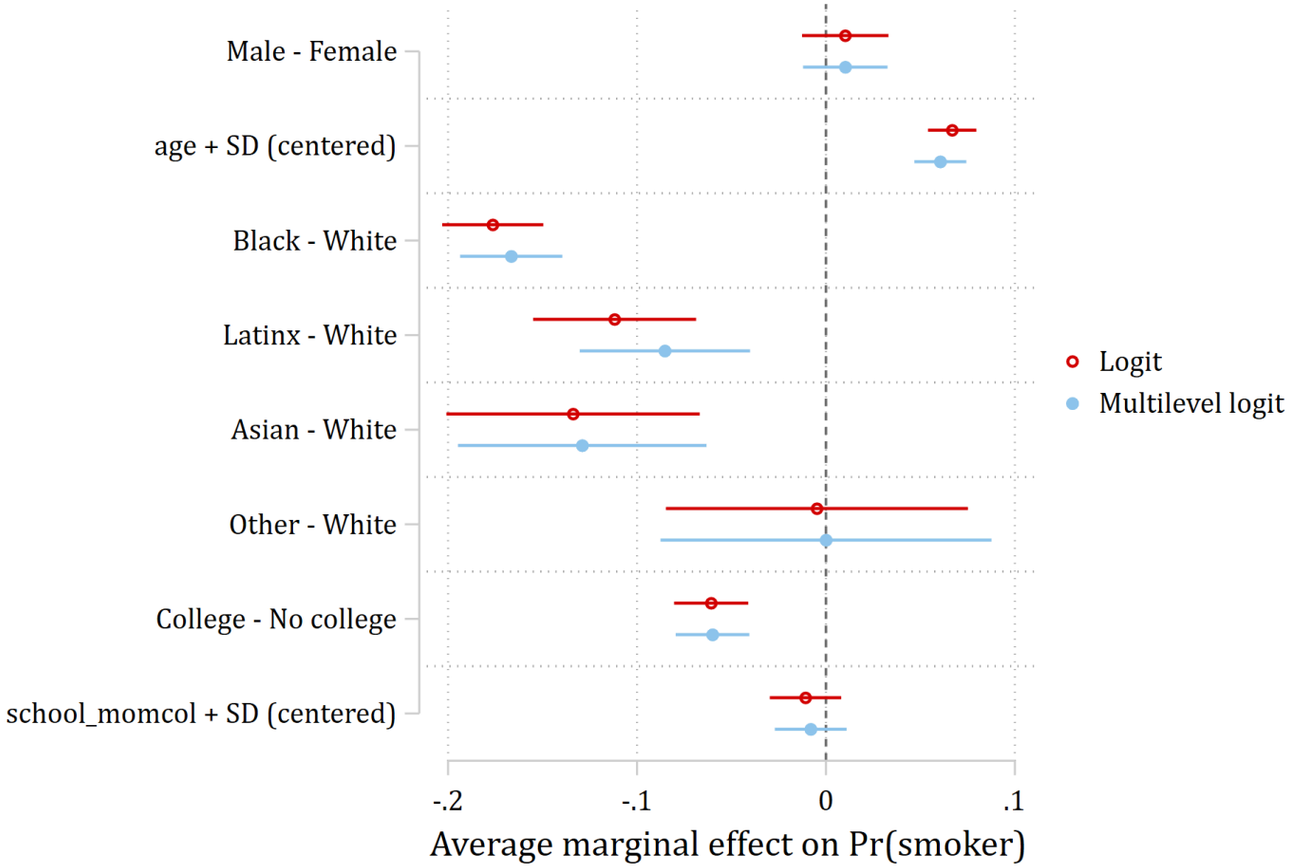
1. Correlated random effects: add each person's mean of the time-varying predictors

```
bysort hhidpn: egen mean_marstatB = mean(marstatB)
bysort hhidpn: egen mean_age = mean(age)
quietly xtlogit depressed i.marstatB c.age i.woman i.race4cat i.collegeB, re intpoints(30)
estimates store re
quietly xtlogit depressed i.marstatB c.mean_marstatB c.age c.mean_age ///
    i.woman i.race4cat i.collegeB, re intpoints(30)
estimates store cre
mecompare marstatB, models(re cre)
```

```
                               |  ME #   Estimate  Robust SE      P>|z|
-------------------------------+-----------------------------------------
marstatB                       |
       Married - Not married   |
                           re  |     1     -0.090      0.003      0.000
                          cre  |     2     -0.059      0.006      0.000
                   Difference  |     3     -0.032      0.005      0.000
```

Within persons, being married lowers the probability of depression by 0.059, about a third less than the random-effects estimate.

Example: multilevel models



1. Add Health Wave 1: adolescents nested in schools; smoking
2. An ordinary logit vs. a multilevel logit with a random intercept for school
3. The two models give similar effects; the age effect and the Latinx–White gap are a bit smaller in the multilevel model

Multilevel models: tests

```
quietly melogit smoker i.male c.age i.race i.momcollege c.school_momcol || school:
estimates store multilevel
quietly logit smoker i.male c.age i.race i.momcollege c.school_momcol
estimates store ordinary
mecompare, models(ordinary multilevel) amount(sd) store(ml)
```

```
                                 |  ME #   Estimate  Robust SE      P>|z|
---------------------------------+-----------------------------------------
age + SD (centered)              |
                         ordinary |     4      0.067      0.007      0.000
                       multilevel |     5      0.061      0.007      0.000
                       Difference |     6      0.006      0.003      0.013
---------------------------------+-----------------------------------------
race                             |
                Black - White    |
                         ordinary |     7     -0.176      0.014      0.000
                       multilevel |     8     -0.167      0.014      0.000
                       Difference |     9     -0.010      0.005      0.042
              Latinx - White     |
                         ordinary |    10     -0.112      0.022      0.000
                       multilevel |    11     -0.085      0.023      0.000
                       Difference |    12     -0.027      0.009      0.004
```

Selected rows. No `vce(robust)`: `mecompare` clusters the standard errors on school.

Most effects are similar, but the age effect and the Black–White and Latinx–White gaps are significantly smaller in the multilevel model.

`totalme` across models

```
quietly mlogit healthR i.college i.race4 c.age i.woman i.parent i.married, vce(robust)
estimates store basemod
quietly mlogit healthR i.college i.race4 c.age i.woman i.parent i.married ///
    c.faminc, vce(robust)
estimates store medmod
totalme college, models(basemod medmod)
```

```
Total ME Estimates (N_basemod = 19292 , N_medmod = 19292)

                         |   Estimate   Std. err.           z       P>|z|
-------------------------+----------------------------------------------
college                  |
       Model 1 (basemod) |      0.159       0.006      26.338       0.000
        Model 2 (medmod) |      0.118       0.007      17.514       0.000
       Cross-Model Diff. |      0.041       0.003      15.006       0.000
```

A college degree moves 0.159 of probability across the health categories; family income accounts for about a quarter of this (0.041).

The SUEST formula

1. Robust (sandwich) variance for each model m

$$\widehat{\mathrm{Var}}\left(\hat{\boldsymbol{\beta}}_m\right) = \mathbf{D}_m^{-1}\left(\sum_i w_{mi}\,\mathbf{u}_{mi}\,\mathbf{u}_{mi}'\right)\mathbf{D}_m^{-1}$$

2. Cross-model covariance for models p and q

$$\widehat{\mathrm{Cov}}\left(\hat{\boldsymbol{\beta}}_p, \hat{\boldsymbol{\beta}}_q\right) = \mathbf{D}_p^{-1}\left(\sum_i w_i\,\mathbf{u}_{pi}\,\mathbf{u}_{qi}'\right)\mathbf{D}_q^{-1}$$

3. $\mathbf{D}_m$: the Jacobian of model m; $\mathbf{u}_{mi}$: the equation-level scores for observation i
    a. Linear regression: the residual; maximum likelihood: the first derivative of the log likelihood
4. Each observation is treated as a cluster

suest2 in practice

```
quietly logit union i.collgrad c.age i.south, vce(robust)
estimates store m1
quietly logit union i.collgrad c.age i.south c.wage c.ttl_exp, vce(robust)
estimates store m2
suest2 m1 m2
```

1. One set of results: equations `m1_union` and `m2_union`, with every covariance
2. Then `test`, `lincom`, `nlcom`, `predict`, and `margins` work within and across models
3. Supports most models, plus survey data, weights, and multiple imputation
4. `mecompare` runs `suest2` for you

Custom tests with `metest`

```
metest 1 - 2                // difference
metest (1 - 2) / 1          // proportion change
metest 1 / 2                // ratio
metest (1 + 2) / #2         // average of two effects
metest 1 = 2 = 3            // joint test of equality
metest (1 = 2) (4 = 5)      // two constraints at once
metest abs(1) - abs(2)      // functions: abs(), min(), exp(), ln(), ...
metest 1 - 2, add           // build a table of tests
```

1. Numbers refer to the rows of the last `mecompare` table
2. Also works after `margins, post` and other estimation commands

What the new commands can do

Compared with existing software, the commands work for:

1. Single or multiple models
2. Single-level, multilevel, and longitudinal models
3. Any amount of change for continuous predictors
4. Summary measures for nominal and ordinal variables
5. Comparisons across variables on different metrics
6. Comparisons across different model types
7. Custom and nonstandard tests of effects

Supported models

1. **Ordinary:** `regress`, `logit`, `probit`, `ologit`, `oprobit`, `mlogit`, `poisson`, `nbreg`, `zip`, `zinb`, `glm`, `cloglog`, `tobit`, `intreg`, `heckman`, `streg`, `gologit2`, `ivregress 2sls`, `fracreg`, `betareg`, `truncreg`, `hetprobit`, `biprobit`, `ivprobit`, `ivtobit`
2. **Panel:** `xtreg` (re, fe, be, mle, pa), `xtlogit` (re, fe, pa), `xtprobit` (re, pa), `xtologit`, `xtoprobit`, `xtmlogit` (re, fe), `xtpoisson` (re, fe, pa), `xtnbreg` (re, pa), `xtcloglog` (re, pa)
3. **Multilevel:** `mixed`, `melogit`, `meprobit`, `mecloglog`, `mepoisson`, `menbreg`, `meologit`, `meoprobit`, `mestreg`, and `meglm` (Gaussian-identity and Gamma-log)
4. Multilevel and panel models can be combined with `regress`, `logit`, `probit`, `cloglog`, `poisson`, `nbreg`, `ologit`, or `oprobit`; all must share the highest-level group

Survey data, multiple imputation, and weights

1. **Survey data:** fit each model with `svy:`; linearized standard errors
2. **Probability weights:** `regress`, `logit`, `probit`, `poisson`, `nbreg`, `ologit`, `oprobit`, `mlogit`
3. **Multiple imputation:** fit each model with `mi estimate, post:`
    a. The models are estimated jointly within each imputation and pooled with Rubin's rules
    b. Ordinary, `xtreg` (mle), and most multilevel models
4. Then `mecompare` as usual