Fungible Regression Coefficients

2024 Portland Stata Conference

Phil Ender - UCLA (Ret)*

August 2024

*Actual affiliation: LRCC (Living room, Culver City)

Prologue

Back in 2010 Robert MacCallum (UNC Chapel Hill) gave a presentation to our departmental colloquium series on fungible parameter weights in structural equation modeling. In his talk he mentioned that fungible weights also applied to multiple regression.

His presentation piqued my interest so I decided to look deeper into this topic.

Webster says…

fun•gi•ble

Adjective

(of a product or commodity) replaceable by another identical item; freely exchangeable for or replaceable by another of like nature or kind; mutually interchangeable: *money is fungible — money that is raised for one purpose can easily be used for another.*

It is a truism that cash is fungible.

It is also true that mushrooms are not fungible. They are fungi.

Ordinary Least Squares Regression (OLS)

- The goal in OLS regression is to estimate a set of weights (coefficients) such that the residual sum of squares (RSS) is a minimum.

- Further, the R-square between the response variable and the linear combination of the predictors ($r^2_{Y\hat{Y}}$) is a maximum, which I will denote as R^2_{max}.

- Thus, $(1-R^2_{max})$•TotalSS = RSS

- There is one, and only one, set of such weights (coefficients) that meet this requirement, i.e., the solution is unique.

But, what if …

- But what if you estimate an alternative set of coefficients such that R^2_a is only a small fraction less than R^2_{max}? Say, one-half of one percent (0.005) less.

- Let’s call the R-square for the alternative weights R^2_a.

- Thus, $R^2_{max} - R^2_a = R^2_{max} - 0.995R^2_{max} = 0.005$

- How close are these fungible coefficients to the original coefficients? Do they cluster around the observed coefficients? How many fubgible coefficients are possible?

I'll spare you the suspense

- As it turns out, with three or more predictors there are an infinite number of sets of coefficients that yield a difference from R^2_{max} of 0.005.
- Further, these coefficients can, and do, look very different from the original coefficients.
- So, I sorta wrote a program to explore these fungible coefficients.

regfungible.ado

The command **regfungible.ado** will estimate alternative fungible regression weights.

```
regfungible, sets(#) theta(#) prefix(string) seed(#) print

where,
sets(#)        : number of sets of weights -- may be larger than _N
theta(#)       : difference of RSQb-RSQa (default = .01)
prefix(string): prefix for new variables (default is "v_")
seed(#)        : set random seed
print          : display additional output
```

What do I mean by sorta wrote a program

I didn't actually write regfungible as much as I translated it from an R function by Niels G. Waller (2008) from the University of Minnesota.

To be honest, I don't really understand all of the math in his Psychometrika article so I won't try to explain it.

Although I'm not very fluent in R, translating Waller's code into Mata was relatively straight forward.

Example using hsbdemo.dta dataset

```
. use https://stats.idre.ucla.edu/stat/data/hsbdemo, clear

. regress socst read math science
 *   regress required before running regfungible

. regfungible, sets(1000) theta(.005) prefix(w_) seed(19)
```

Note: the number sets can be larger than the number of observations.

Output 1

```
OLS fungible regression weights analysis

 Original R2: RSQb =   .4187076
 Reduced R2:  RSQa =   .4137076
 theta = RSQb-RSQa =       .005
 r_yhata_yhatb     =   .9940113
```

regfungible.ado generated 1,000 sets of coefficients each of which had an R^2_a of .4137076. The correlation between the predicted values from the original OLS and the predicted using the alternative coefficients was .9940113.

Output 2

```
Standardized OLS regression weights

              1              2              3
 1    .4480658545    .2199795336    .0440053932
```

These are the standardized coefficients from the original OLS regression. Note the low coefficient for variable 3.

The fungible regression weights are displayed as standardized weights for computational ease. regfungible does not compute raw fungible regression weights only the standardized coefficients.

Output 3

Maximum fungible regression weights for each variable

	1	2	3
1	.5309603441	.1535344376	.00027434
2	.3843593917	.3153093707	.0067823647
3	.4030944979	.1795592346	.1406651808

Minimum fungible regression weights for each variable

	1	2	3
1	.3544721562	.2806466729	.08729649
2	.500337233	.1193873672	.0811681338
3	.4820592823	.2554627866	-.0537063386

Explanation of Output 3

Here's how to read the tables for Output 3. Of the 1,000 sets of weights the highest coefficient for the first predictor (read) is .5310 (rounded). The other two values in that row are to coefficients for math and science when read takes on its maximum value (observation 25).

The second column and row are interpreted in a similar manner for the variable math. The third column and row for science.

The table for minimum weights works in a similar manner.

Output 4

```
Summary of fungible regression weights

   Stats |       w_1        w_2        w_3
---------+---------------------------------
       N |      1000       1000       1000
    Mean |  .4472829   .2147603   .0404114
      p5 |  .3557416   .1201262  -.0520651
     p25 |  .3821528   .1443142  -.0253707
     p50 |  .4543421   .2134189   .0398007
     p75 |  .5104564   .2843172   .1042769
     p95 |  .5303973   .3135495   .1388187

Observed |  .4408659   .2199795   .0440054
```

Note: The means of the 1,000 fungible weights are, in fact, very close to the observed standardized OLS coefficients.

Generated Fungible Coefficients

The **regfungible** command generated 1,000 sets of regression weights. Each set has an R^2_{alt} of 0.417076 with the original response variable. Here are the first 10 sets of fungible coefficients:

```
. list w_* in 1/10

     +--------------------------------+
     |      w_1         w_2       w_3 |
     |--------------------------------|
  1. | .5303477    .1452587   .010781 |
  2. | .5205218    .1260296  .0464213 |
  3. | .4989157    .1194042  .0830479 |
  4. | .5145808    .2085491 -.0422424 |
  5. | .3685594    .3112611  .0326337 |
     |--------------------------------|
  6. | .4852165    .2517878 -.0536229 |
  7. | .4432466    .2913173 -.0438174 |
  8. | .5028303    .2283259  -.049693 |
  9. | .3613705    .3049778  .0495953 |
 10. | .5292507     .140233  .0181302 |
```

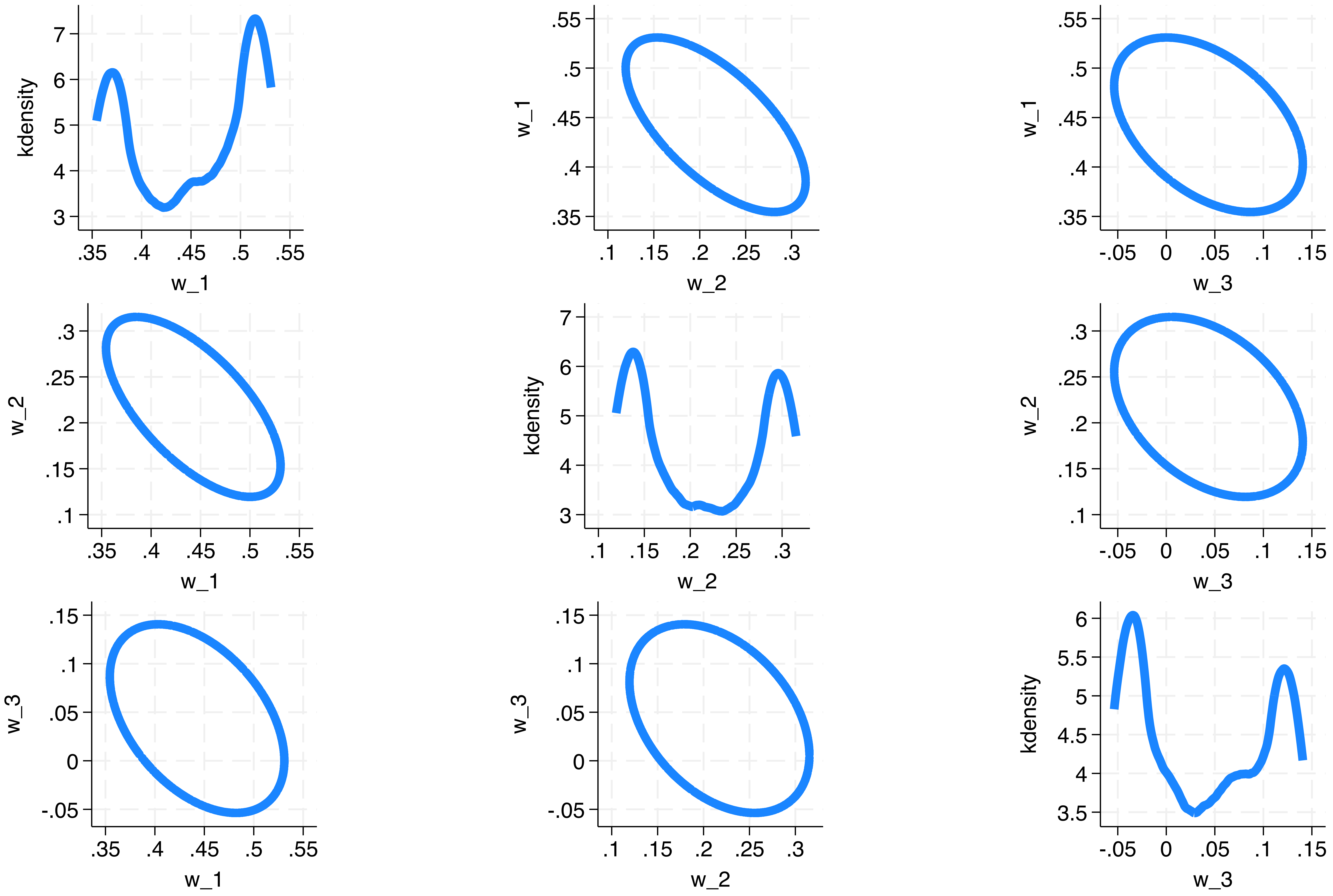
A closer look at observation 25

```
. list w_* in 523/527

     +------------------------------------+
     |       w_1         w_2          w_3 |
     |------------------------------------|
 23. |  .5301372    .1642883    -.0112148 |
 24. |  .3992974    .1840107     .1405279 |
 25. |  .5309603    .1535344     .0002743 |
 26. |  .4292087    .1536423     .1361107 |
 27. |  .3718804    .3128987     .0262792 |
```

Correlations among the fungible coefficients

```
. corr w_*
(obs=1,000)

             |      w_1      w_2      w_3
-------------+---------------------------
         w_1 |   1.0000
         w_2 |  -0.6792   1.0000
         w_3 |  -0.4350  -0.3654   1.0000
```

Matrix graph of fungible coefficients

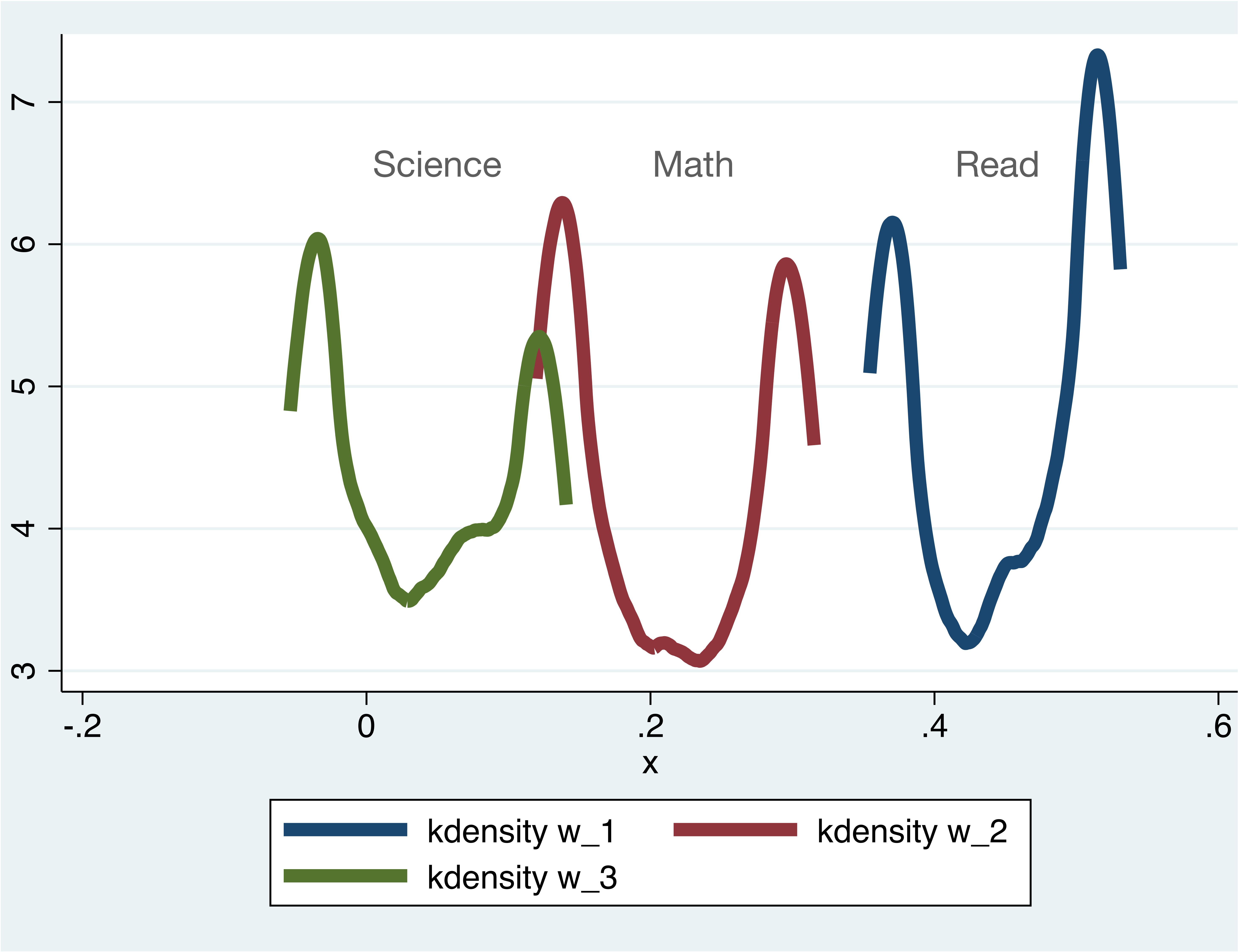


With three predictors the fungible weights would graph onto the surface of an ellipsoid.

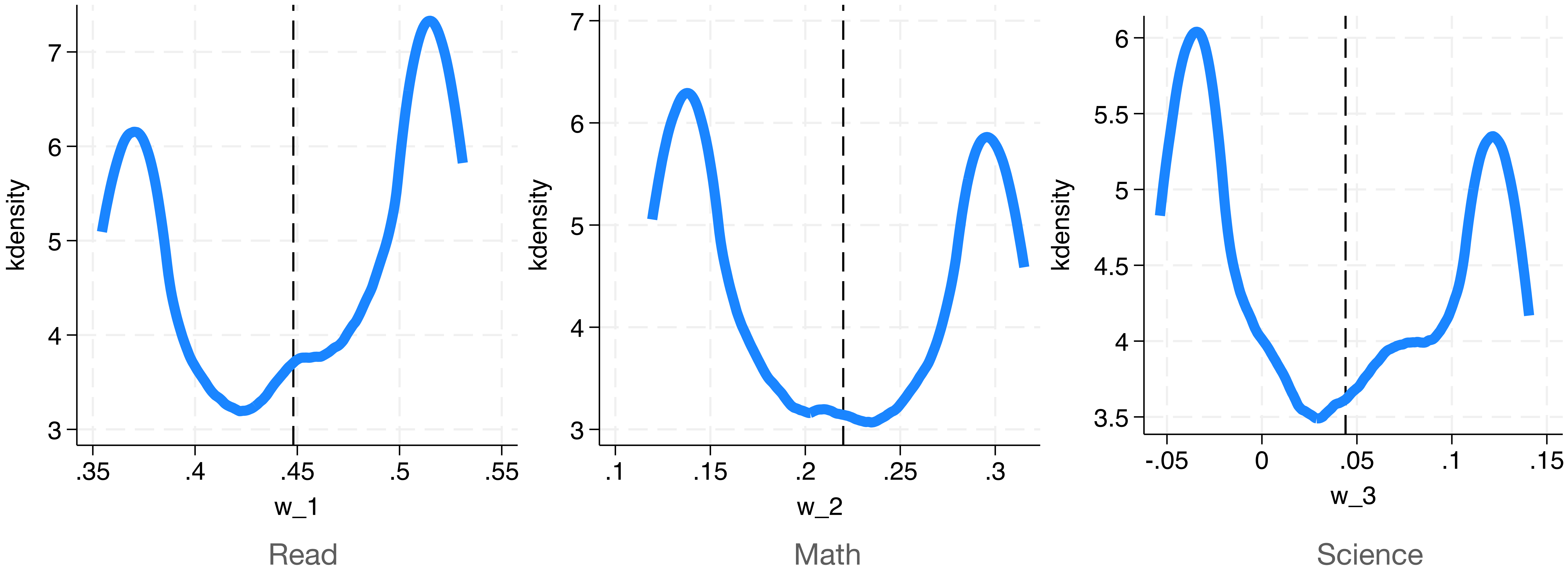
However, we will make do with a scatterplot matrix.

Ellipses are not lines but 1,000 scatterplot points of fungible coefficients.

Kernel density plots of fungible coefficients

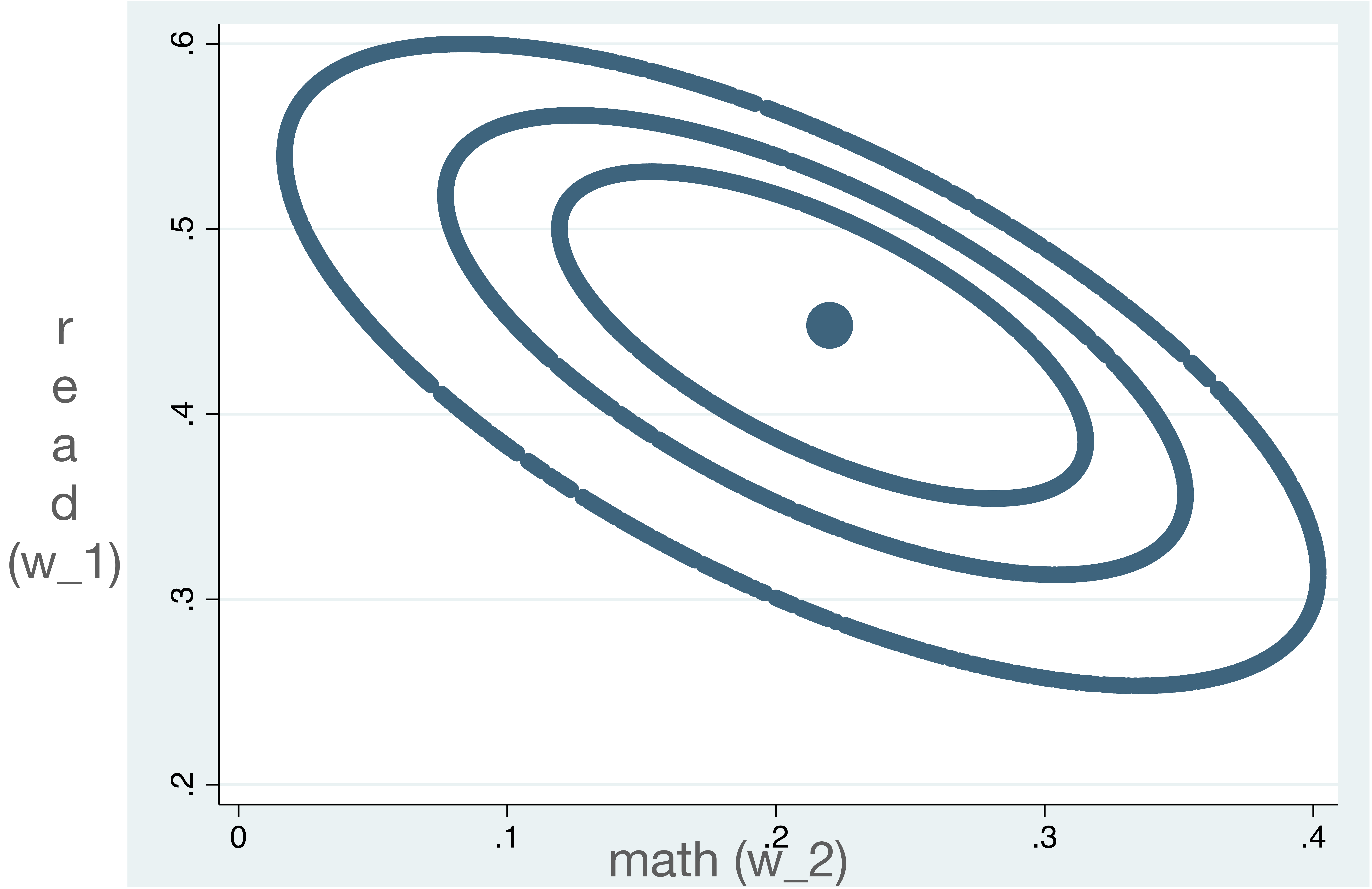


Kernel density plots revisited



Vertical lines at observed standardized coefficients.

Fungible weights at 2%, 1%, & 0.5%



What about predicted scores?

I generated three predicted standard scores (zhat1, zhat2, zhat3) using coefficients the from the table of maximum fungible regression weights for each variable. Recall…

	1	2	3
1	.5309603441	.1535344376	.00027434
2	.3843593917	.3153093707	.0067823647
3	.4030944979	.1795592346	.1406651808

```
. sum zhat1 zhat2 zhat3
```

Variable	Obs	Mean	Std. dev.	Min	Max
zhat1	200	-1.70e-09	.643201	-1.413063	1.417385
zhat2	200	2.48e-09	.643201	-1.238322	1.475447
zhat3	200	-1.74e-09	.643201	-1.264543	1.45036

Note: All 3 have the same standard deviation but different min and max. All have R^2 of .413707 with the response variable (Not shown).

Predicted standardized values

```
. list zhat1 zhat2 zhat3, clean

           zhat1       zhat2       zhat3
  1.   -1.135539   -1.090984   -1.264543
  2.   -1.135345   -1.086189   -1.165092
  3.   -.8275256   -.8046313   -1.053091
  4.   -.9636809   -.9421247   -1.070602
  5.   -.8927192   -.9303521   -.9450608
  6.   -.7048054    -.756056   -.9024416
  7.   -1.208678   -1.028315   -1.144583
  8.   -.3232099   -.5214121   -.5836322
  9.   -1.007355   -1.163893   -1.036604
 10.   -1.053319   -.9158516   -1.026638
 11.    -1.15165   -1.117791   -1.141636
 12.   -.9252744   -.9921849    -.869735
 13.   -.4129591   -.4978791   -.5964971
 14.    -1.09716    -1.14173   -1.077883
 15.   -1.099531   -1.072083    -.931833
```

Why does regfungible generate standardized coefficients?

Mathematically and computationally it’s much easier to estimate the standardized fungible coefficients.

However, it is fairly straight forward to covert standardized coefficients to raw coefficients.

Converting standardized coefficients to raw regression coefficients.

$$b_{xi}=B_{xi}*SD_y/SD_{xi}$$

Fungible weights are interesting, but are they useful?

Waller (2008) suggests that fungible regression weights are useful as a kind of sensitivity analysis providing an alternative method of estimating parameter variability (an alternative to bootstrap or likelihood methods).

I like to show fungible coefficients to students when they turn in their first multiple regression projects. Students have a tendency to believe that the coefficients in their model are a window into the “truth”. That fungible coefficients can fit the model almost as well comes as a bit of a shock to them.

References

MacCallum, R. 2010. Fungible parameter estimates: Troublesome implications for regression, structural equation modeling, mediation analysis, and more. Presentation at UCLA Graduate School of Education.

Waller, Niels G. 2008. Fungible weights in multiple regression. *Pyschometrika* 73: 691-703.