Applications of Stata's 'margins' in social science

Philip S. Morrison.
Victoria University of Wellington

Philip.Morrison@vuw.ac.nz

Oceania Stata User Group Meeting.
Australian National University. University House
24-25 September 2015

All Stata commands are in red. Slides are best viewed as a Powerpoint due to animation.

Why 'margins'?

'*Margins*' is a term used to describe a suite of post-estimation steps that can be applied to most Stata estimation procedures, e.g. regress, logit etc.

It generates adjusted means (predictive margins) immediately after running any permissible model.

It offers convenient ways of contrasting adjusted means (predictions) to judge the statistical influence of values of categorical and continuous arguments (and their interactions).

These useful features are enhanced by the complementary graphics (*marginsplot*).

The social sciences has appeared slow to adopt the use of 'margins'.

In this brief talk I hope to convince you of its value in research and teaching using three cases studies of: inequality preferences, loneliness, and job satisfaction.

Some references

Jann, B. (2014). Plotting regression coefficients and other estimates. *The Stata Journal, 14*(4), 708-737.

Mitchell, M. N. (2012). *Interpreting and visualizing regression models using Stata*. College Station, Texas: A Stata Press Publication

Royston, P. (2013). Marginscontplot: plotting the marginal effects of continuous predictors. *The Stata Journal, 13*(3), 510-527.

Stata. (2013). Margins, marginsplot, contrasts etc..

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal, 12*(2), 308-331.

Williams, R. (2013). Using Stata's margins command to estimate and interpret adjusted predictions and marginal effects. (Powerpoint)

Example 1. Inequality preferences: does age matter?

Our attitudes to income distribution have important behavioural and political consequences.

New Zealand views are both very conservative and heterogeneous

In this example I ask whether peoples age is related to their inequality preference

Data: the 2004 World Value Survey. N = 900 +/-

Steps

a) Regress responses to a question income inequality preferences on age (OLS)

b) Generate 'adjusted means' via margins

c) Graphically display the relationship between the adjusted means and age categories via marginsplot

c) Test differences in adjusted means via contrast.

d) Contrast means across age groups in three ways: reference, adjacent, Helmet

Table 1. Preferences for income *in*equality.
World Values Survey (2004): New Zealand.

```
.tab e035 if s025a  == 5542004 // New Zealand

                    income equality |      Freq.     Percent        Cum.
------------------------------------+-----------------------------------
  1. incomes should be made more equal |        106       11.78       11.78
                               2. 2 |         49        5.44       17.22
                               3. 3 |         83        9.22       26.44
                               4. 4 |         74        8.22       34.67
                               5. 5 |        126       14.00       48.67
                               6. 6 |         95       10.56       59.22
                               7. 7 |        147       16.33       75.56
                               8. 8 |        131       14.56       90.11
                               9. 9 |         31        3.44       93.56
10. we need larger income differences a |         58        6.44      100.00
------------------------------------+-----------------------------------
                              Total |        900      100.00

. sum e035 if s025a  == 5542004 // New Zealand

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        e035 |       900    5.427778    2.623227          1         10
```

The model:

$$(1) \qquad I_i = \beta_o + \beta_1 A_{25<35} + \ldots + \beta_6 A_{65+} + \varepsilon_i$$

Table 3. Regression of income inequality responses on age indicators. World Values Survey, New Zealand 2004

```
. regress inequal i.age6cat if s025a  == 5542004 // New Zealand

      Source |       SS       df       MS              Number of obs =     881
-------------+------------------------------           F(  5,   875) =    2.46
       Model |  83.4075331     5  16.6815066           Prob > F      =  0.0320
    Residual |     5943.16   875  6.79218286           R-squared     =  0.0138
-------------+------------------------------           Adj R-squared =  0.0082
       Total |  6026.56754   880   6.8483722           Root MSE      =  2.6062

------------------------------------------------------------------------------
     inequal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     age6cat |
          2  |  -.4538321   .4115824    -1.10   0.270    -1.261636     .353972
          3  |  -.5703279   .3811931    -1.50   0.135    -1.318487    .1778317
          4  |  -.7410215    .388083    -1.91   0.057    -1.502704    .0206608
          5  |  -.9076006   .3905279    -2.32   0.020    -1.674081   -.1411198
          6  |  -1.168207   .3905279    -2.99   0.003    -1.934687   -.4017259
             |
       _cons |   6.180328   .3336874    18.52   0.000     5.525407    6.835249
------------------------------------------------------------------------------
```

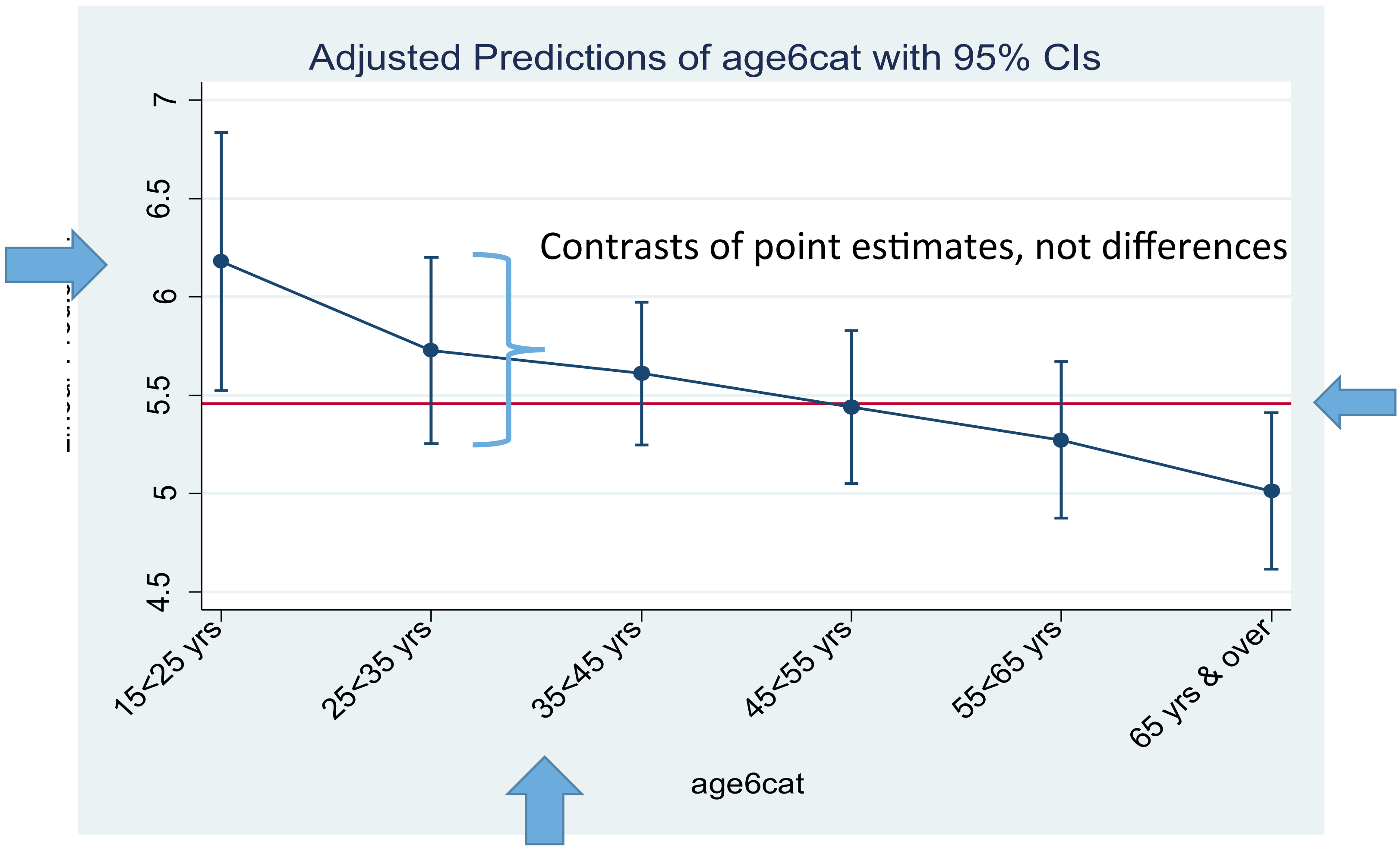
```
display _b[_cons] + _b[1.age6cat] =  6.1803279

display _b[_cons] + _b[2.age6cat] =  5.7264957
```

**Table 4. Predicted margins of income inequality by age.
World Values Survey, New Zealand 2004**

```
. margins i.age6cat

Adjusted predictions                              Number of obs   =        881
Model VCE    : OLS

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     age6cat |
  15<25 yrs  |   6.180328   .3336874    18.52   0.000     5.525407    6.835249
  25<35 yrs  |   5.726496   .2409416    23.77   0.000     5.253605    6.199387
  35<45 yrs  |       5.61   .1842849    30.44   0.000     5.248308    5.971692
  45<55 yrs  |   5.439306   .1981443    27.45   0.000     5.050413      5.8282
  55<65 yrs  |   5.272727    .202891    25.99   0.000     4.874517    5.670937
65 yrs & over|   5.012121    .202891    24.70   0.000     4.613911    5.410331
------------------------------------------------------------------------------
```

Figure 1. The preference for income (in)equality by age. New Zealand 2004

```
.marginsplot, yline(5.456) xlabel (, angle(45))
```



Confidence intervals and contrasts

Table 5. Reference contrasts of adjusted predictions of income inequality preferences by age. New Zealand, 2004.

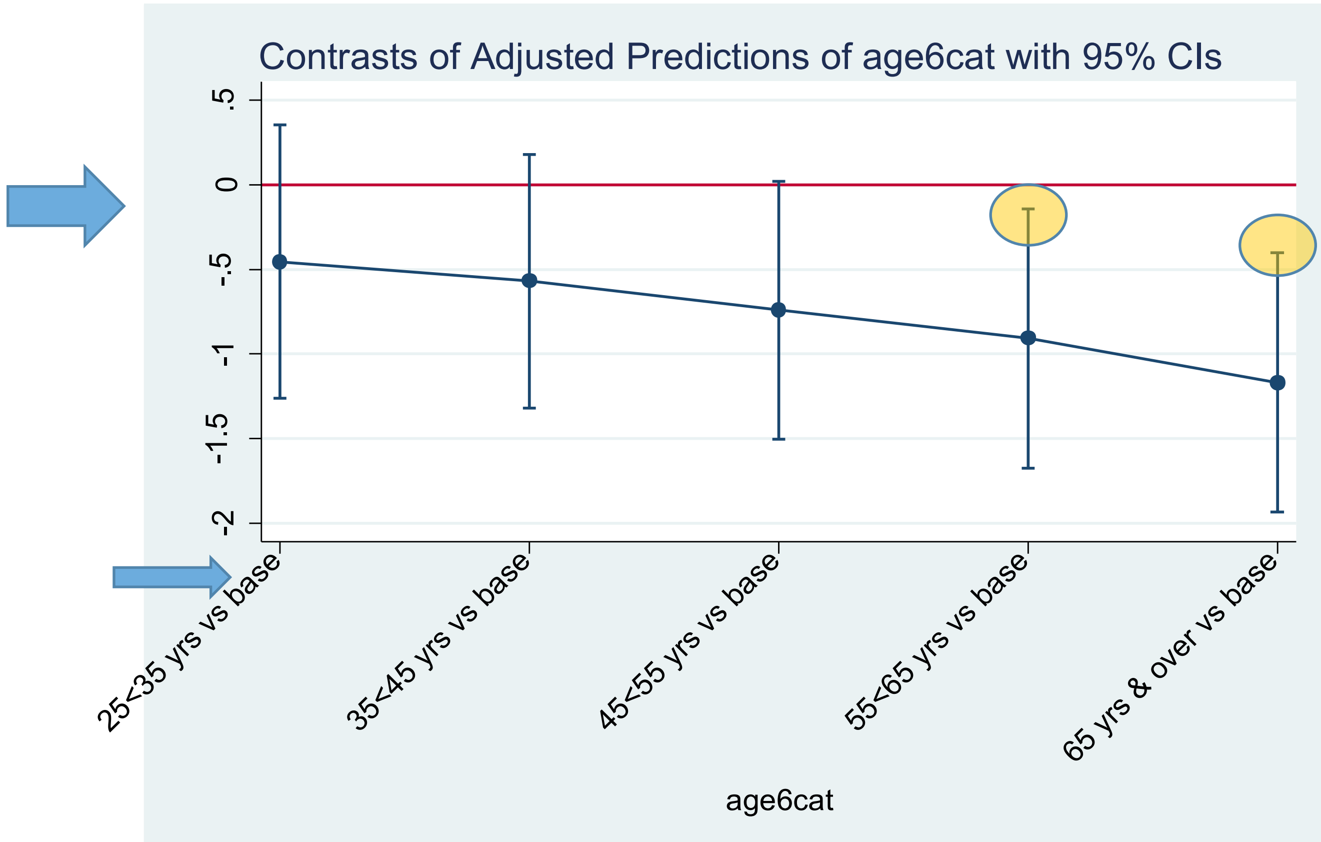
```
. margins i.age6cat, contrast (nowald effects)

Contrasts of adjusted predictions
Model VCE    : OLS

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------------------
                         |            Delta-method
                         |   Contrast   Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
                 age6cat |
    (25<35 yrs vs base)  |  -.4538321   .4115824    -1.10   0.270    -1.261636     .353972
    (35<45 yrs vs base)  |  -.5703279   .3811931    -1.50   0.135    -1.318487    .1778317
    (45<55 yrs vs base)  |  -.7410215    .388083    -1.91   0.057    -1.502704    .0206608
    (55<65 yrs vs base)  |  -.9076006   .3905279    -2.32   0.020    -1.674081   -.1411198
(65 yrs & over vs base)  |  -1.168207   .3905279    -2.99   0.003    -1.934687   -.4017259
------------------------------------------------------------------------------------------
```

Alternative to generate similar output…

```
.contrast r.age6cat, (nowald effects)
```

Figure 2. Contrasting adjusted predictions of preference for income inequality of each age group against the base. New Zealand 2004.

```
.marginsplot, yline(0) xlabel (, angle(45))
```



Point:
Older NZers favour greater equality

Table 6. Adjacent contrasts of adjusted predictions of income inequality preferences by age. New Zealand, 2004.

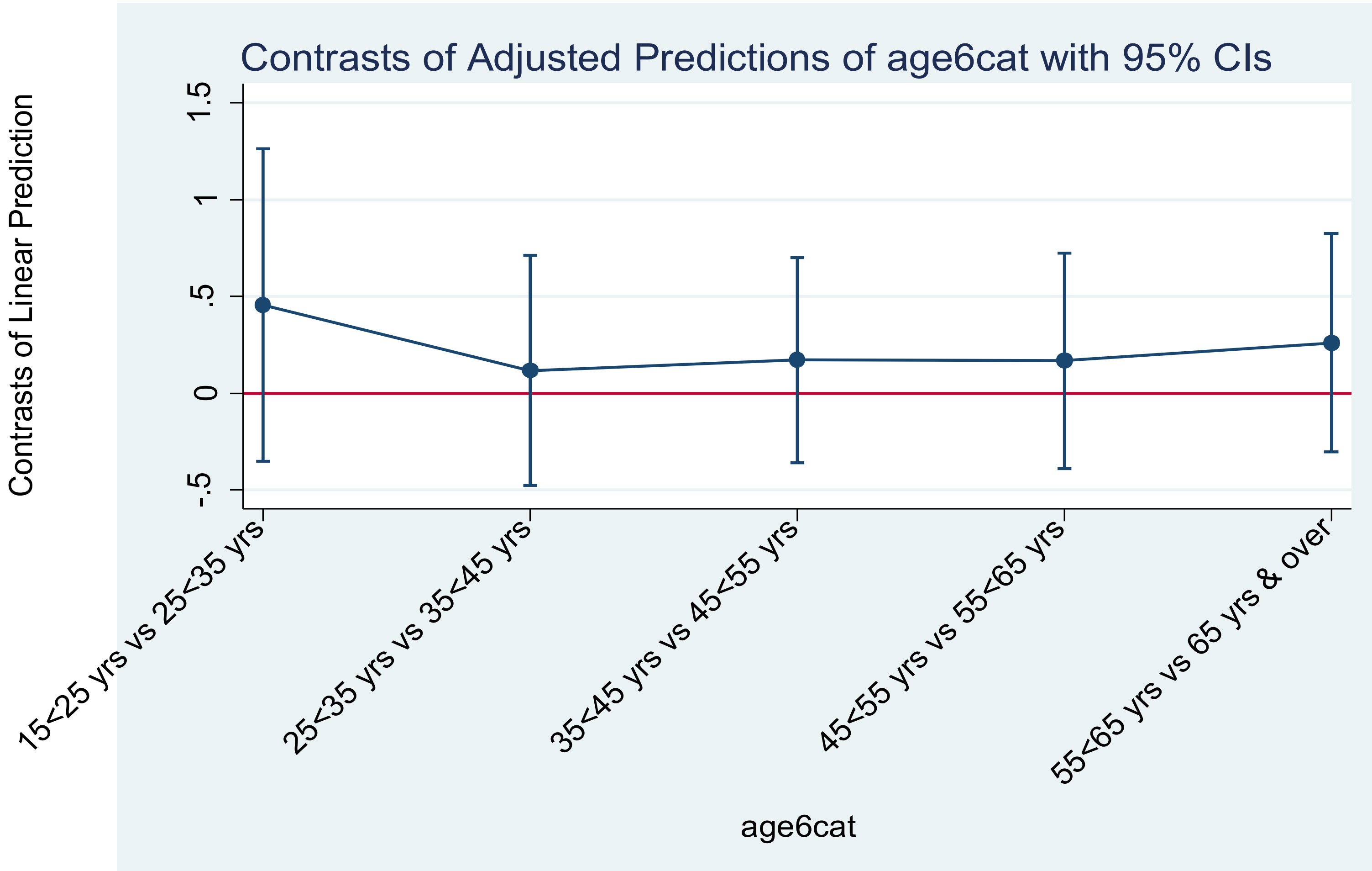
```
. margins a.age6cat, contrast (nowald effects)

Contrasts of adjusted predictions
Model VCE    : OLS

Expression   : Linear prediction, predict()
```

age6cat	Contrast	Delta-method Std. Err.	t	P>|t|	[95% Conf.	Interval]
(15<25 yrs vs 25<35 yrs)	.4538321	.4115824	1.10	0.270	-.353972	1.261636
(25<35 yrs vs 35<45 yrs)	.1164957	.3033377	0.38	0.701	-.4788588	.7118502
(35<45 yrs vs 45<55 yrs)	.1706936	.2705958	0.63	0.528	-.360399	.7017863
(45<55 yrs vs 55<65 yrs)	.1665791	.2835946	0.59	0.557	-.3900261	.7231843
(55<65 yrs vs 65 yrs & over)	.2606061	.2869312	0.91	0.364	-.3025477	.8237598

Neighbouring age groups exhibit no statistical difference in preference for (in)equality

Figure 3. Contrasting the adjacent margins (adjusted predictions) of preference for income inequality by age group. New Zealand 2004

```
.marginsplot, yline(0) xlabel (, angle(45))
```



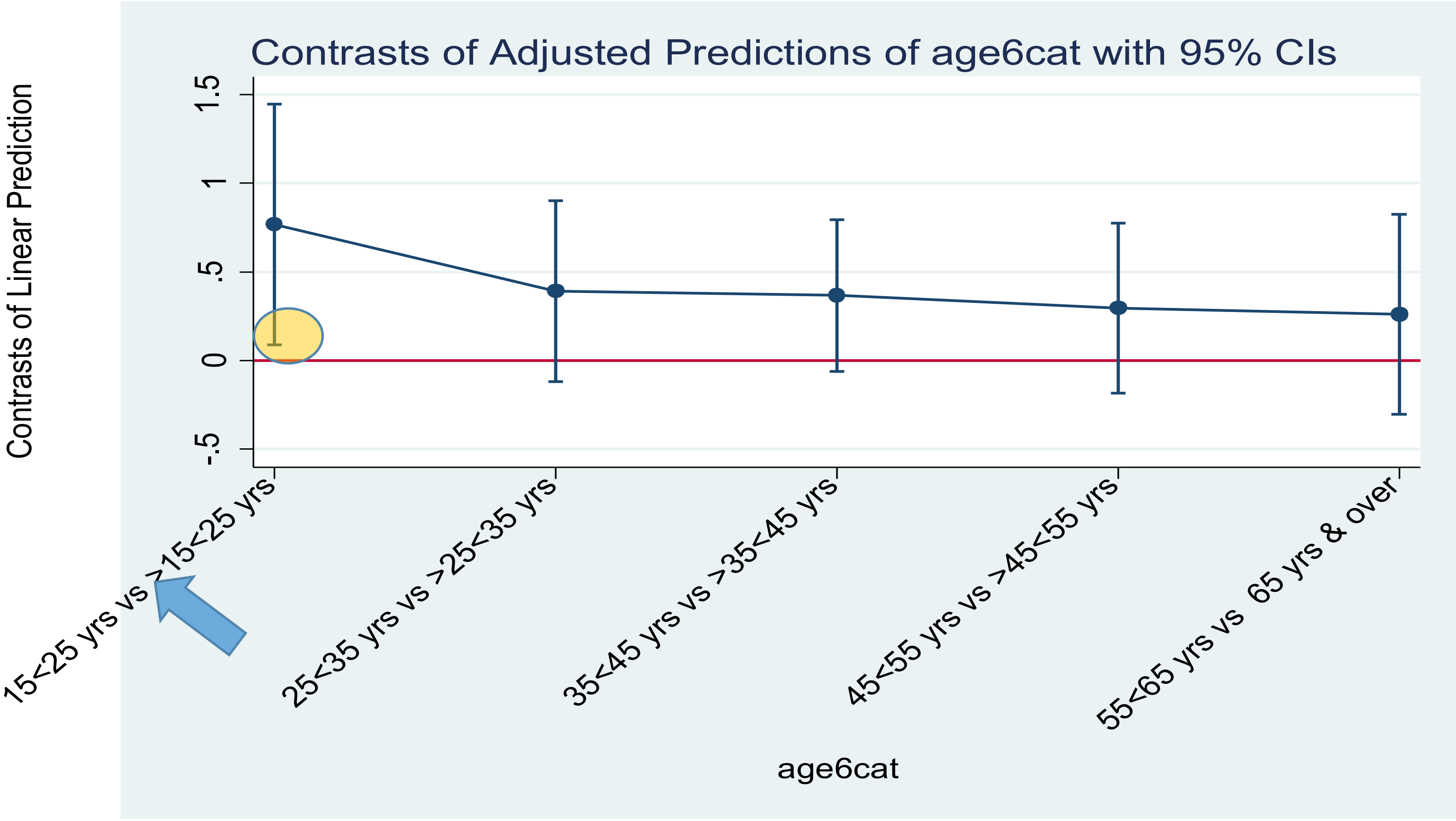
Helmert contrasts

Contrasts each age group with the mean of those following. Identifies thresholds

```
.margins h.age6cat, contrast(nowald pveffects)

Contrasts of adjusted predictions
Model VCE    : OLS

Expression   : Linear prediction, predict()

-----------------------------------------------------------------------
                               |          Delta-method
                               |   Contrast   Std. Err.      t    P>|t|
-------------------------------+---------------------------------------
                       age6cat |
    (15<25 yrs vs >15<25 yrs)  |   .7681978   .3462534     2.22   0.027
    (25<35 yrs vs >25<35 yrs)  |    .392957    .260336     1.51   0.132
    (35<45 yrs vs >35<45 yrs)  |   .3686151   .2178784     1.69   0.091
    (45<55 yrs vs >45<55 yrs)  |   .2968821   .2446294     1.21   0.225
(55<65 yrs vs  65 yrs & over)  |   .2606061   .2869312     0.91   0.364
-----------------------------------------------------------------------
```

Figure 4. Helmert contrasts. Differences in the margins of preference for income inequality at different thresholds from the youngest age group upwards. New Zealand 2004

```
.marginsplot, yline(0) xlabel(,angle(45))
```



In summary

Method

1. 'Margins' is a suite of post-estimation commands
2. It enables the user to test propositions about predicted values (margins)
3. In particular it facilitates the contrast of one prediction against another
4. It graphically displays predicted values and confidence intervals
5. And, in the case of contrasts, the confidence intervals of user defined difference
6. Covariates? Not included here be see examples below.

Substance

1. New Zealanders views on (in)equality are very heterogeneous
2. Age of the respondent is negatively correlated with preferences for greater inequality

Example 2. Loneliness: the effects of social connection.

Loneliness shortens life and reduces quality of life.

Aim: to model loneliness as a function of social connectivity to illustrate the non-linear case.

Data: New Zealand General Social Survey, 2012. Confidentialised Unit Record File (CURF).

The following:

a) Applies margins to a *logistic* regression model

b) Highlight differences in metrics

c) Tests interaction effects

d) Logistic model applied is as follows

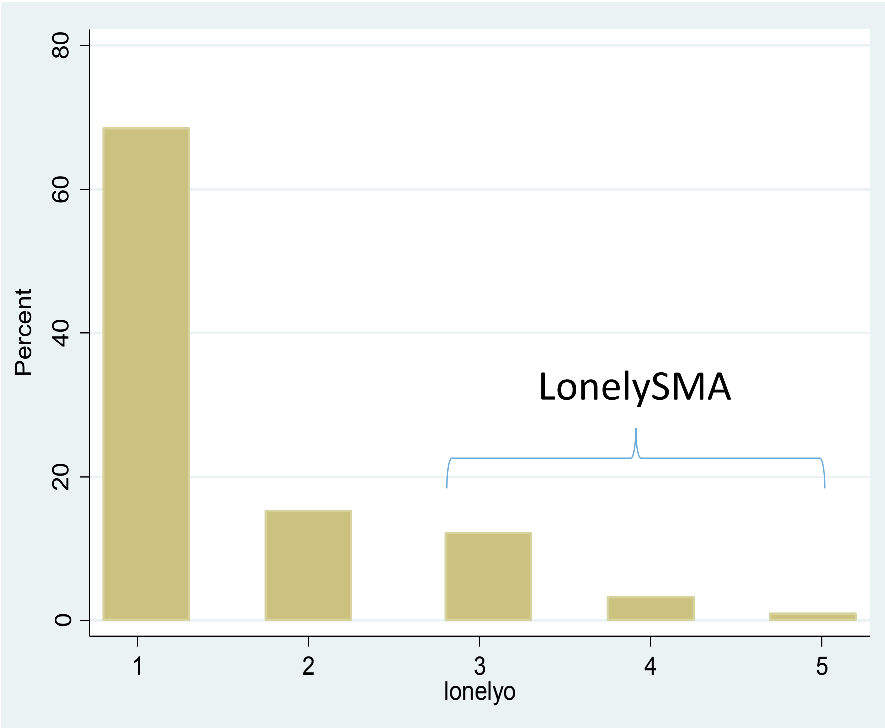
$$(1) \quad L_i = \beta_o + \sum_{j=1}^{J} \beta_j C_{ij} + \varepsilon_i$$

where L_i is a binary loneliness measure and C_j is the type of contact reported in the last week/month. [Covariates omitted].

Statistics New Zealand disclaimer:

"Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the author, not Statistics New Zealand."

Q: *In the past four weeks, how often have you felt isolated from others?*

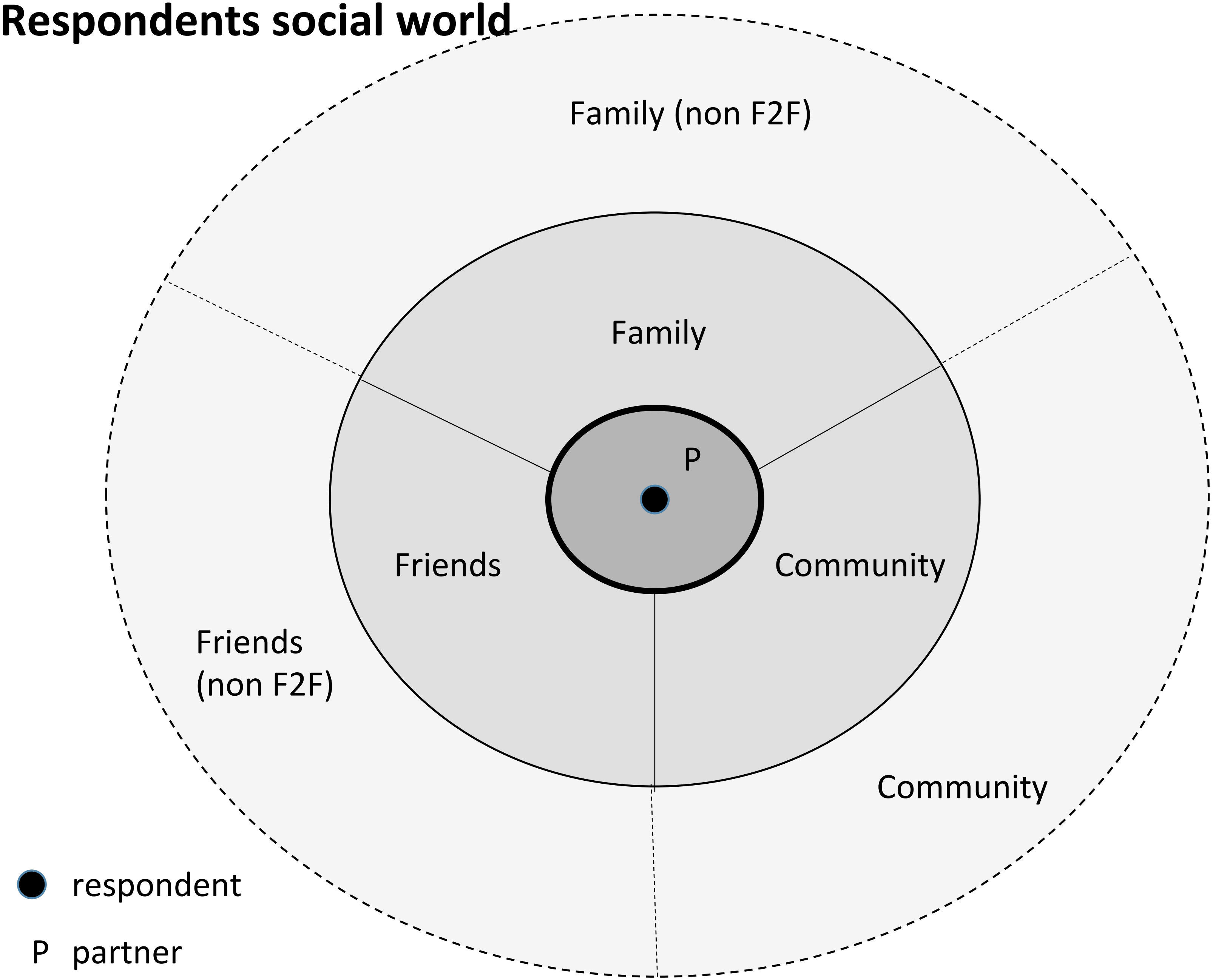


LonelySMA = 16%

lonelyo	Freq.	Percent	Cum.	Population	
1.None of the time	5,415	68.45	68.45	3.03	
2.A little of the time	1,202	15.19	83.64	0.67	
3.Some of the time	963	12.17	95.82	0.54	
4.Most of the time	258	3.26	99.08	0.14	
5.All of the time	73	0.92	100.00	0.04	
Total	7,911	100.00		4.43 mill.	LonelySMA = 0.72 mill

Source: New Zealand General Social Survey, 2012

Respondents social world



Just over half of all respondents have partners living in the household (58.2%), nearly 85% have had contact with family living locally over the past month, and over 92 % have had contact with local friends.

Begin by exploring 2 x 2 interaction effects on loneliness

Table 1. Empirical probability of loneliness by partnership and local family.

```
. table partner family, c(mean lonelySMA count lonelySMA) format(%9.3g)

--------------------------------------
           |         family
   partner | No family      Family
-----------+--------------------------
No-Partner |      .258  >     .195  = prob. of lonelySMA
           |       569       2,732  = counts
           |
   Partner |      .211  >     .121
           |       635       3,975
--------------------------------------
```

Table 2. The 2 x 2 case. The influence of partner and family on loneliness in the log-odds metric. New Zealand, 2012

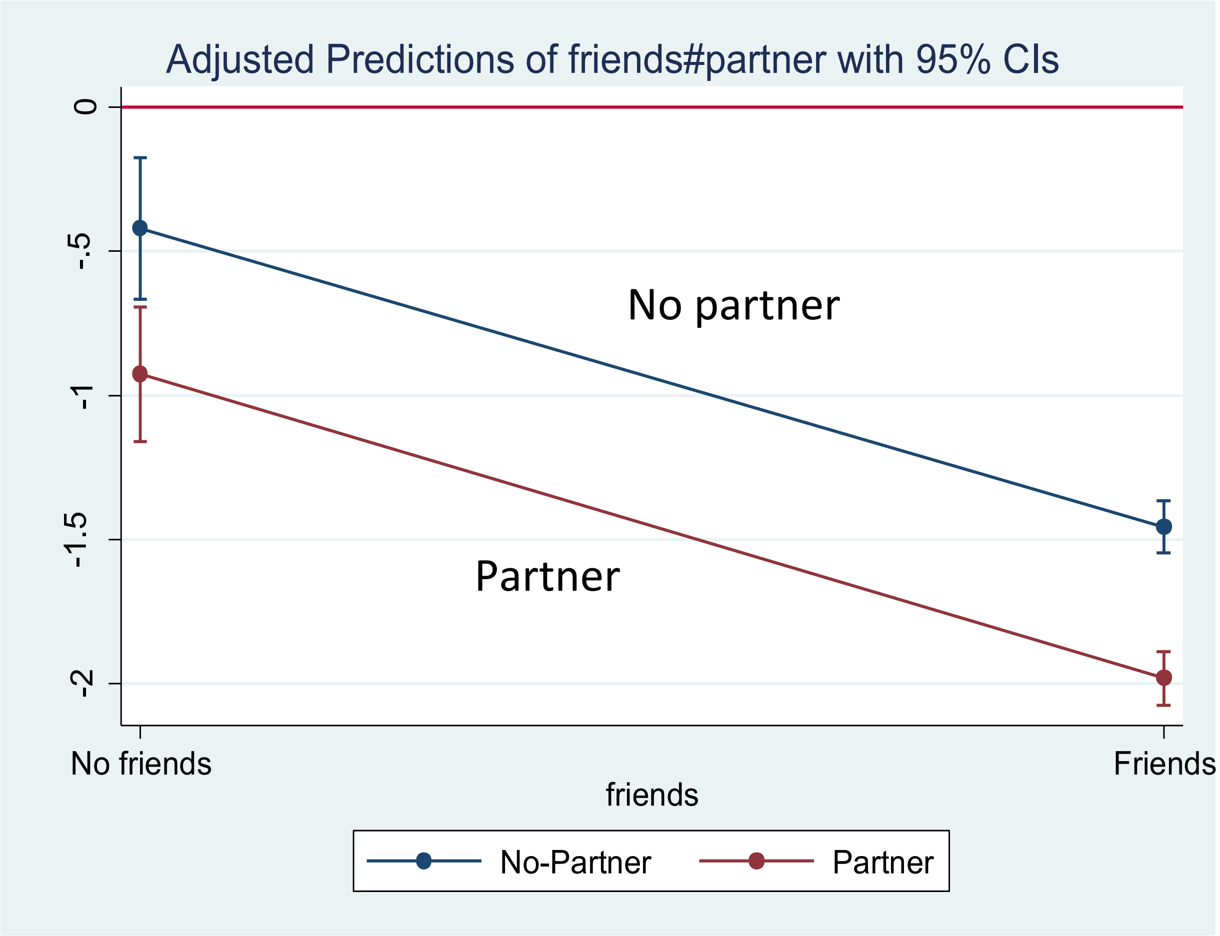
```
. logistic lonelySMA i.family##partner

Logistic regression                               Number of obs   =       7911
                                                  LR chi2(3)      =     118.01
                                                  Prob > chi2     =     0.0000
Log likelihood = -3465.6794                       Pseudo R2       =     0.0167

------------------------------------------------------------------------------
      lonelySMA | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
         family |
      No family |          1  (base)*
         Family |   .6941993   .0744661    -3.40   0.001     .5625701    .8566269
                |
        partner |
     No-Partner |          1  (base)
        Partner |   .7678251   .1048048    -1.94   0.053     .5875934    1.003339
                |
 family#partner |
 Family#Partner |   .7414313   .1132487    -1.96   0.050     .5496111    1.000199
                |
          _cons |   .3483411   .0333615   -11.01   0.000     .2887241    .4202682
------------------------------------------------------------------------------

*set showbaselevels on, permanently
```

Table 3. The 2 x 2 case. The influence of partner and family on loneliness in the *log-odds* metric. New Zealand, 2012

```
. margins friends#partner, predict(xb)

Adjusted predictions                              Number of obs   =       7910
Model VCE    : OIM

Expression   : Linear prediction (log odds), predict(xb)

------------------------------------------------------------------------------------------
                       |            Delta-method
                       |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------+------------------------------------------------------------------
       friends#partner |
No friends#No-Partner  |  -.4212135   .1255938    -3.35   0.001    -.6673728   -.1750541
   No friends#Partner  |  -.9263411   .1187477    -7.80   0.000    -1.159082   -.6935999
   Friends#No-Partner  |    -1.4561   .0463501   -31.42   0.000    -1.546944   -1.365255
      Friends#Partner  |  -1.981803   .0469583   -42.20   0.000     -2.07384   -1.889767
------------------------------------------------------------------------------------------
```

Figure 1. Margins in the (linear) log-odds metric. The parallel impact of partners and friends on log-odds of being lonely (SMA). New Zealand, 2012

```
.marginsplot, yline(0)
```



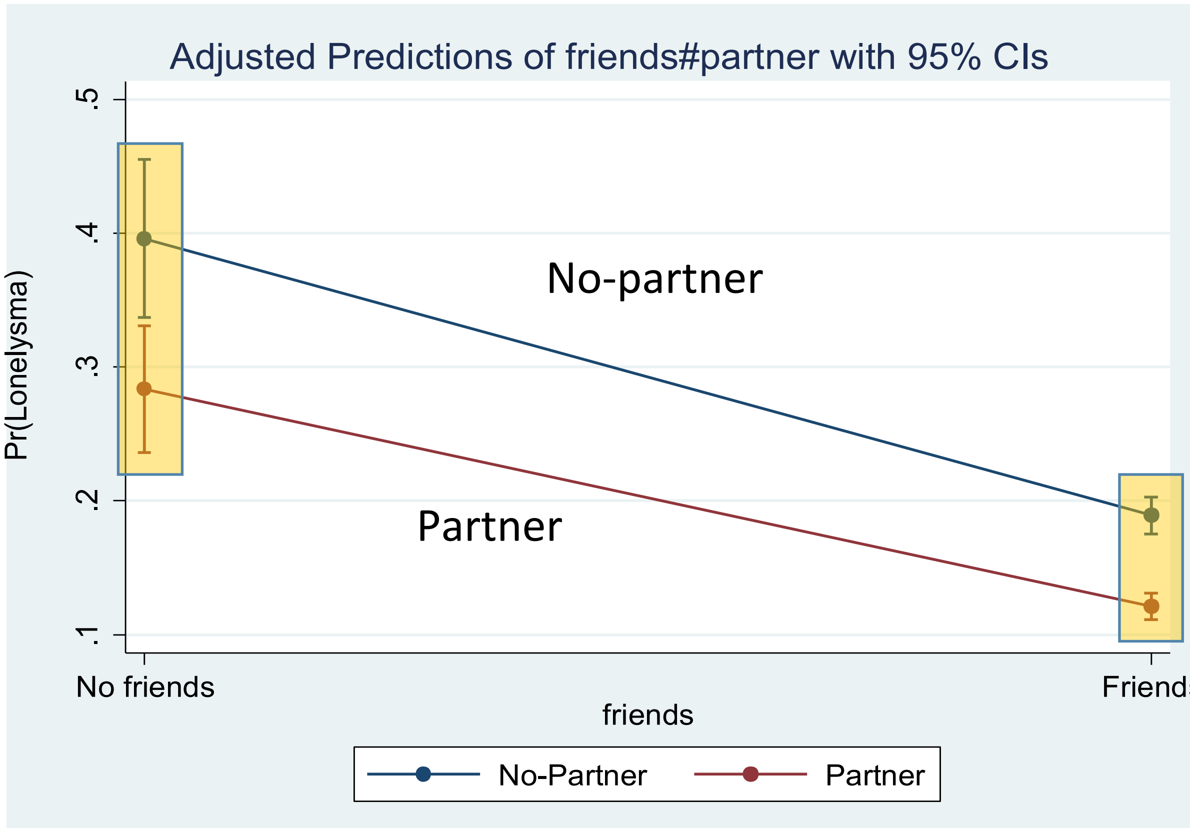
Shows not interaction in the log-odds metric

Table 2. The 2 x 2 case. The influence of partner and family on loneliness in the *probability* metric. New Zealand, 2012

```
. margins friends#partner

Adjusted predictions                              Number of obs   =       7910
Model VCE    : OIM

Expression   : Pr(lonelySMA), predict()

------------------------------------------------------------------------------------------
                      |            Delta-method
                      |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------+-------------------------------------------------------------------
      friends#partner |
No friends#No-Partner |   .3962264   .0300459    13.19   0.000     .3373375    .4551154
   No friends#Partner |   .2836676   .0241296    11.76   0.000     .2363745    .3309607
   Friends#No-Partner |   .1890646   .0071064    26.60   0.000     .1751363    .2029928
      Friends#Partner |   .1211268   .0049989    24.23   0.000      .111329    .1309245
------------------------------------------------------------------------------------------
```

Figure 2. Margins in the (non-linear) *probability* metric. The impact of partners and friends on probability of being lonely (SMA). New Zealand, 2012

```
.marginsplot
```



The results show an interaction in the probability metric. I.e. there is a difference in the *probability* [cf. log odds] of being lonely (SMA) between those with partners and those without but this difference *diminishes* in the presence of friends. This relationship could be altered in the presence of covariates of course.

In summary

Method

1. 'Margins' reflects the metric. In the non-linear case: log-odds, odds or probability
2. The nature of the interaction also reflects the metric
3. Marginsplot after margins helps interpret interaction effects
4. Non-linear models mean that relationships like these can change as the values of other variables in the model change.

Substance

1. Loneliness falls with social contact
2. Not all types of social contact have the same effect (in cross-section)
3. The joint presence of different types of contact (e.g. partner + friends) can reduce loneliness over and above their separate effects
4. How this interaction effect operates may depend on age, income, education etc. and the suite of margins commands allows such hypotheses to be tested explicitly

Example 3. Job satisfaction: the effects of job insecurity.

Job satisfaction is very sensitive to job insecurity but the interaction between the two is poorly understood.

Aim: to model job satisfaction as a function of job security [+/- covariates].

a) In this example I combine margins with Jann's Coefplot

Statistics New Zealand disclaimer:

"Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the author, not Statistics New Zealand."

Figure 1. The estimated relationship between job satisfaction (S) and job insecurity (I)

(2) $S\downarrow i = \beta\downarrow o + \sum j{=}1 \uparrow J ▒ \beta\downarrow ij\ I\downarrow i + \varepsilon\downarrow i$
$= S\downarrow i = \beta\downarrow o + \beta\downarrow 1\ I\downarrow low + \beta\downarrow 1\ I\downarrow med + \beta\downarrow 1\ I\downarrow high + \beta\downarrow 1\ I\downarrow certain +$

Figure 2. Estimated marginal effects of job insecurity on average job satisfaction under the full set of controls. New Zealand 2008+2012. (Using coefplot)

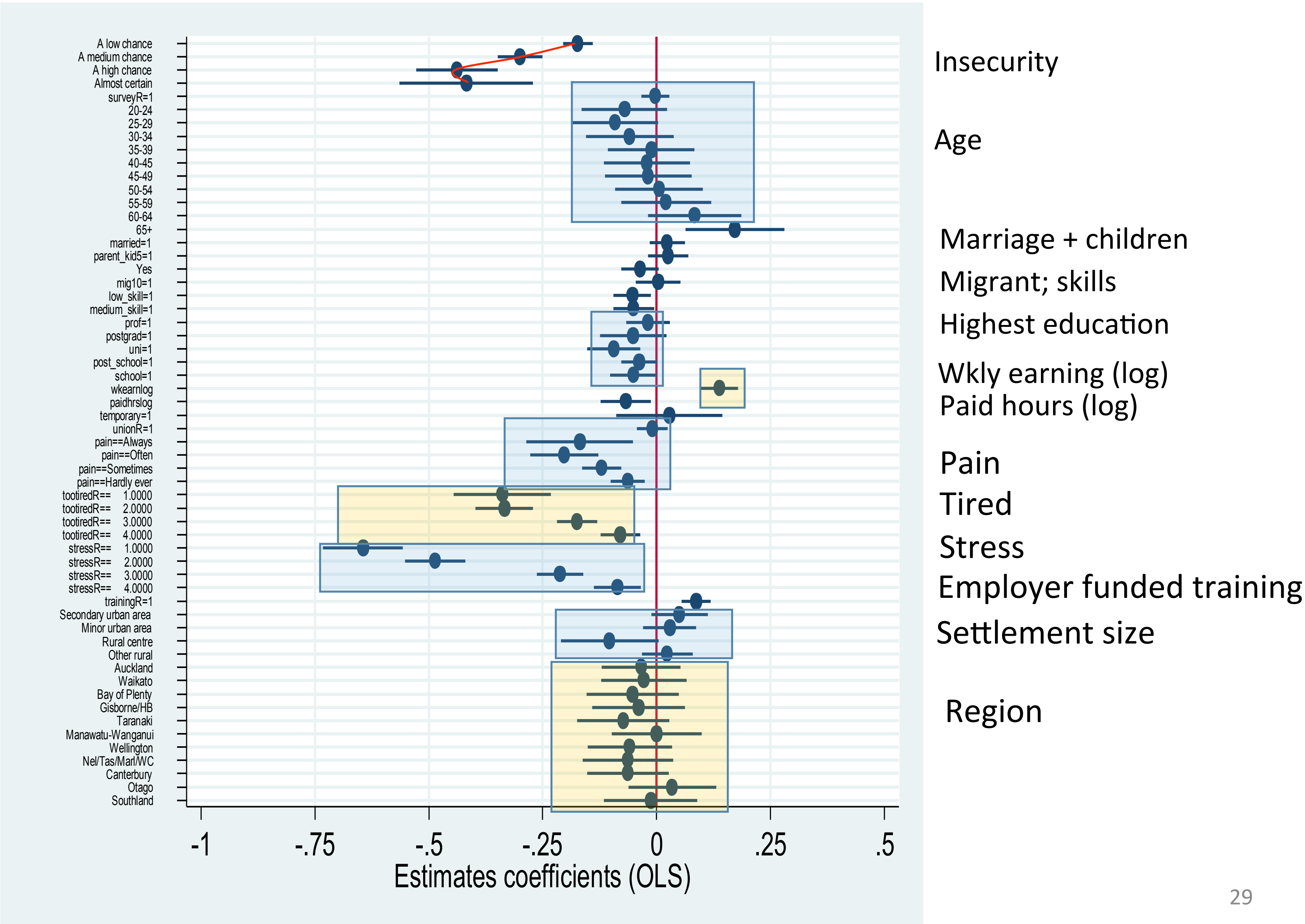
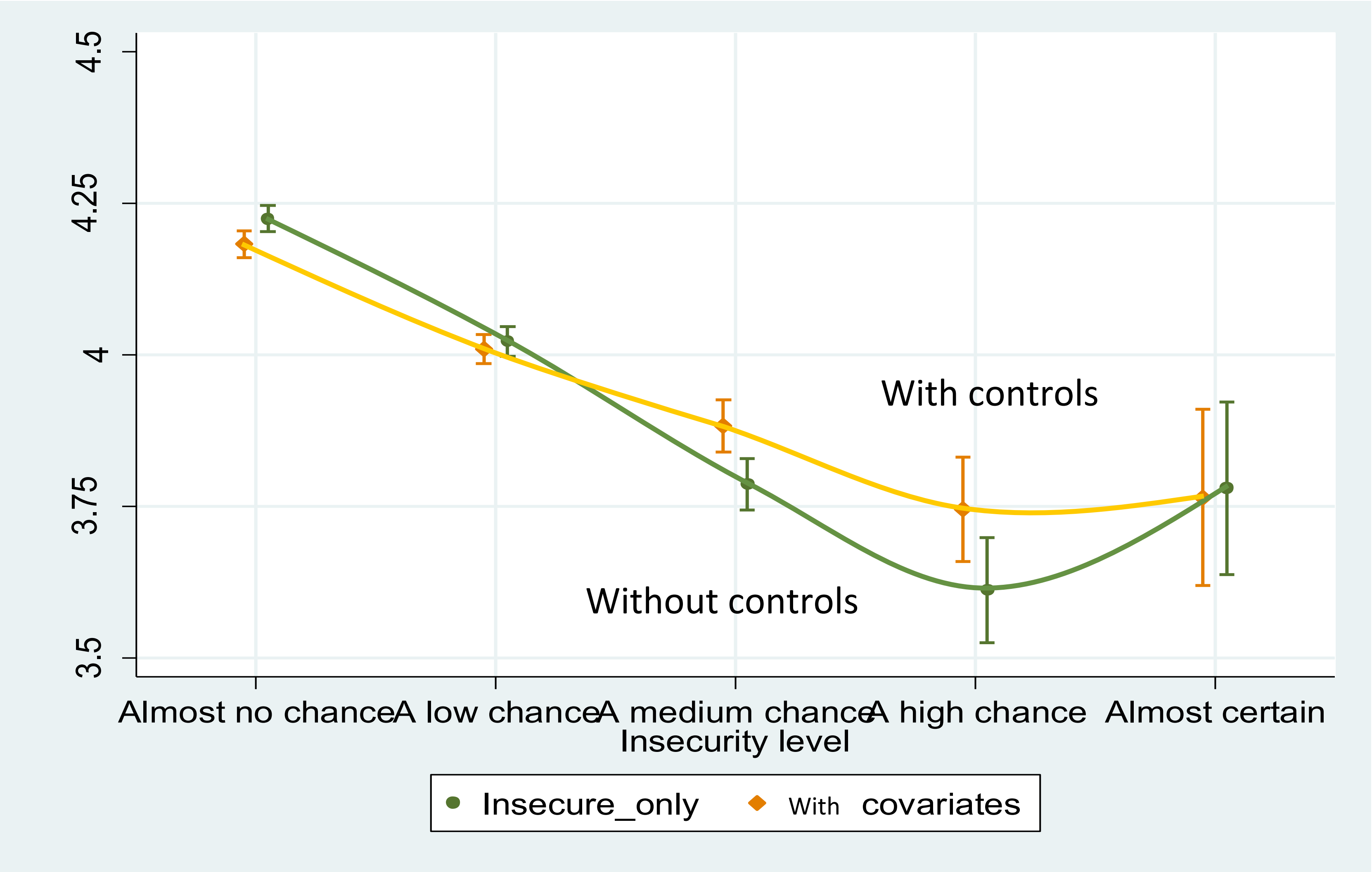


Figure 3. Estimated average levels of job satisfaction by level of job insecurity before and after controls. Male permanent employees. New Zealand 2008 and 2012



Controls flatten (reduce) the influence of insecurity on satisfaction - slightly

QUESTIONS?