# Annex – supplementary results.

This annex includes the main model results from the full set of models run during the Fragile Families Challenge, and which were not part of the main manuscript. It also shows the model of layoff that was best-placed at the mid-point of the competition, and hence was eligible for the relevant 'Progress Prize'.

Models.

GPA

As with the models of grit and material hardship, this was based on a random forest. The 25 most important variables (48 were included in the model) are shown in Figure 1. The variables which had the greatest effect on the prediction were linked to academic ability, such as scores on the Woodcock Johnson tests, and teachers' views about children's mathematical skills. The prediction from the model of material hardship was also somewhat important, to a much greater degree than the prediction from the model of grit, and a constructed measure of grit based on wave 5 information. Measures of parental education were also somewhat important in terms of the predictions made of the child's GPA.

Figure 1   Variable importance measures for GPA model (Final Stage –ranked 31st on holdout data).
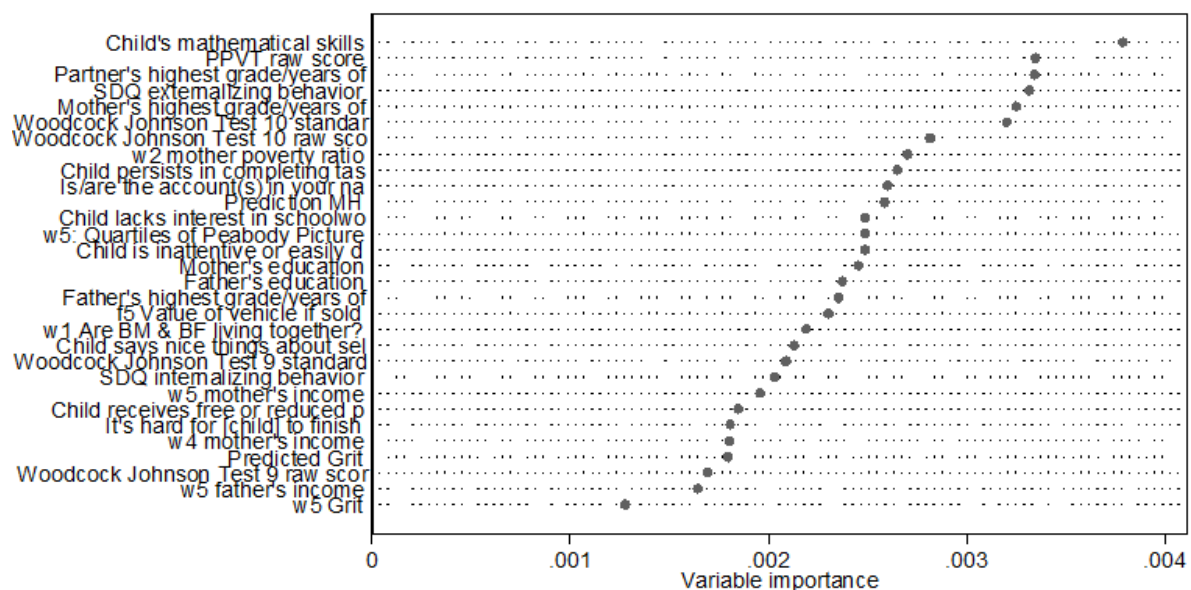
Grit

See main paper for details.

Material hardship

See main paper for details.

Eviction

The model of eviction used a binary logistic regression model, with results set out in Table 1. There appeared to be a strong association with a constructed measure of material hardship from wave 5. Evictions were also associated with being behind on the rent or mortgage (at wave 5) and with the level of rent being paid at wave 5. An increasing level of the mother's income (at wave 5) was also associated with a lower risk of eviction by wave 6.

In relative terms this was the poorest model I produced, ranking at 108th and doing worse than a simple mean prediction.

Table 1 Logistic regression model of eviction (rank = 108)

Variables	Coefficients (standard error)
Mother's education level 2	0.753***
	(0.287)
Mother's education level 3	0.019
	(0.366)
Mother's education level 3	1.102**
	(0.523)
Material hardship = 1	0.672
	(0.443)
Material hardship = 2	0.588
	(0.500)
Material hardship = 3	1.758***
	(0.433)
Material hardship = 4	1.661***
	(0.488)
Material hardship = 5	1.925***
	(0.522)
Material hardship = 6	2.384***
	(0.575)
Material hardship = 7	2.710***
	(0.689)
Material hardship = 8	2.989***
	(0.996)
Material hardship = 9	-9.554

Variables	Coefficients (standard error)
	(535.411)
Material hardship = 10	2.903*
	(1.509)
Mother's wave 1 income	-0.00001**
	(0.00001)
Behind with rent or mortgage	0.634**
	(0.293)
Number of moves since child was 1	0.044
	(0.029)
Level of rent	0.001**
	(0.0003)
Constant	-4.099***
	(0.403)
Observations	1,459
Log Likelihood	-273.782
Akaike Inf. Crit.	583.564

Layoff

The final model is reported in the main paper.

The model that was leading at the mid-point of the Challenge, and hence winning the Progress Prize was a binary logistic regression model, as follows (Table 2). The variables that were significantly associated with the risk of layoffs included the mother and father incomes, and the mother's level of educational attainment. Respondents identifying as Black, non-Hispanic had higher risks of job layoff even accounting for the other factors listed.

Table 2 Logistic regression model of layoffs [mid-point of competition].

Variables	Coefficients (standard error)
Wave 5 material hardship	0.055
	(0.039)
Mother's wave 5 income/10000	-0.064*
	(0.029)
Father's wave 5 income/10000	-0.044+
	(0.024)
Father's wave 5 income is zero	0.639+
	(0.335)
No. of flexible job features	-0.117+
	(0.064)

Variables	Coefficients (standard error)
Race = black, non-Hispanic	0.418*
	(0.206)
Race = Hispanic	0.304
	(0.237)
Female education level 3	0.544**
	(0.203)
Professional occupation	-0.208
	(0.221)
Executive occupation	0.410
	(0.254)

Job training

The binary logistic regression model for job training is show as Table 3. Job training seemed to be rather more common for the Black, non-Hispanic population, and was more common for mothers with higher incomes. Where people had previously been on courses to improve skills, subsequent job training was less likely. This is a mixed and surprising set of results. Even so, it was placed 19th in the competition on the holdout data.

Table 3 Logistic regression model of job training (rank = 19).

Variables	Coefficients (standard error)
Material hardship = 1	0.036
	(-0.191)
Material hardship = 2	0.136
	(-0.217)
Material hardship = 3	0.508**
	(-0.238)
Material hardship = 4	0.542**
	(-0.261)
Material hardship = 5	0.549
	(-0.336)
Material hardship = 6	0.419
	(-0.426)
Material hardship = 7	1.296**
	(-0.522)
Material hardship = 8	0.620
	(-0.916)
Material hardship = 9	-11.405
	(-535.411)
Material hardship = 10	-12.406
	(-352.546)
Race = black, non-Hispanic	0.527***

Variables	Coefficients (standard error)
	(0.180)
Race = Hispanic	0.165
	(0.211)
Have taken classes to improve job skills = YES [w5]	-0.608**
	(0.142)
Have taken classes to improve job skills = SKIPPED [w5]	-0.420
	(0.438)
Attending vocational/technical/trade school = YES	-0.717
	(0.351)
Attending vocational/technical/trade school = SKIP	-0.865
	(0.168)
Mother's wave 5 income/10000	0.099**
	(0.026)
Mother's wave 1 income/10000	0.00000
	(0.00000)
Constant	-1.062***
	(0.303)
Observations	1,461
Log Likelihood	-742.324
Akaike Inf. Crit.	1,522.648