Appendix

Joe Backer, Elle Dodd, Dan Humphrey, Dan Peckham

3/25/2017

```r
library(haven)
library(sandwich)
library(stargazer)
library(ggplot2)
library(car)
library(knitr)
library(plyr)
library(lfe)
library(plm)
library(gtools)
```

```r
afghan <- read_dta("~/Documents/Stats2/pivotproject/afghandata.dta")
afghan <- rename(afghan, c("f07_hh_id"= "hh_id",
                 "f07_heads_child_cnt" = "heads_child",
                 "f07_girl_cnt" = "girl",
                 "f07_age_head_cnt" ="age_head",
                 "f07_yrs_ed_head_cnt" = "yrs_ed_head",
                 "f07_jeribs_cnt" = "jeribs",
                 "f07_num_sheep_cnt" = "num_sheep",
                 "f07_duration_village_cnt" = "duration_village",
                 "f07_farsi_cnt" = "farsi",
                 "f07_tajik_cnt" = "tajik",
                 "f07_farmer_cnt"    = "farmer",
                 "f07_num_ppl_hh_cnt"   = "num_ppl_hh",
                 "f07_test_observed" = "test_observed",
                 "f07_formal_school" = "formal_school",
                 "f07_nearest_scl" = "nearest_scl",

                 # non-matching new names#
                 "f07_age_cnt" = "age_child",
                 "f07_both_norma_total" = "test_score_normalized"
))
afghan$sheep_per_hh_member = afghan$num_sheep / afghan$num_ppl_hh
attach(afghan)
```

```r
# create data frame of only the variables of interest
remove <- c("hh_id", "observation_id")
varlist <- colnames(afghan[, !names(afghan) %in% remove])
balance_variables <- afghan[, !colnames(afghan) %in% remove]

# generate counts
n_treatment <- apply(balance_variables[balance_variables$treatment == 1, ],
    2, function(x) length(which(!is.na(x))))
n_control <- apply(balance_variables[balance_variables$treatment == 0, ], 2,
    function(x) length(which(!is.na(x))))

# generate table
```

```r
balancetable <- cbind(n_control, n_treatment)
# drop treatment row
balancetable <- balancetable[!rownames(balancetable) == "treatment", ]

# run t.tests, skipping treatment[14]
balance_tests <- lapply(varlist[c(1:13, 15:20)], function(x) {
    t.test(as.formula(paste(x, "treatment", sep = "~")), data = balance_variables,
        alternative = "two.sided", mu = 0, paired = FALSE, var.equal = FALSE,
        conf.level = 0.95)
})

# extract and adjust p vals
balance_test_pvals <- t(sapply(balance_tests, function(x) {
    c(x$estimate[], `diff in means` = unname(x$estimate[1]) - unname(x$estimate[2]),
        ci.lower = x$conf.int[1], ci.upper = x$conf.int[2], p.value = x$p.value,
        adj.p.value = p.adjust(x$p.value, method = "bonferroni", n = length(x)),
        x$parameter)
}))

balance_test_pvals <- data.frame(balance_test_pvals, stringsAsFactors = FALSE)
balance_test_pvals[] <- lapply(balance_test_pvals, function(x) as.numeric(as.character(x)))
balancetable <- cbind(balancetable, balance_test_pvals)
balancetable <- round(balancetable, 3)
kable(balancetable)
```

	n_control	n_treatment	mean.in.group.0	mean.in.group.1	diff.in.means	ci.lower	ci.upp
heads_child	730	830	0.911	0.927	-0.016	-0.043	0.0
girl	730	830	0.456	0.478	-0.022	-0.072	0.0
age_child	730	830	8.321	8.322	-0.001	-0.165	0.1
age_head	730	830	40.219	40.090	0.129	-0.991	1.2
yrs_ed_head	730	830	3.101	3.531	-0.431	-0.788	-0.0
jeribs	730	830	1.510	1.498	0.011	-0.211	0.2
num_sheep	730	830	6.404	9.586	-3.181	-4.351	-2.0
duration_village	730	830	27.662	30.172	-2.509	-4.059	-0.9
farsi	730	830	0.205	0.210	-0.004	-0.045	0.0
tajik	730	830	0.204	0.239	-0.034	-0.076	0.0
farmer	730	830	0.729	0.707	0.022	-0.023	0.0
num_ppl_hh	730	830	7.905	8.741	-0.835	-1.163	-0.5
test_observed	730	830	0.925	0.925	-0.001	-0.027	0.0
clustercode	730	830	9.500	7.024	2.476	2.031	2.9
chagcharan	730	830	0.429	0.663	-0.234	-0.282	-0.1
formal_school	730	830	0.264	0.731	-0.467	-0.511	-0.4
nearest_scl	730	830	3.149	2.881	0.268	0.157	0.3
test_score_normalized	675	768	0.006	0.586	-0.580	-0.687	-0.4
sheep_per_hh_member	730	830	0.817	1.142	-0.325	-0.468	-0.1

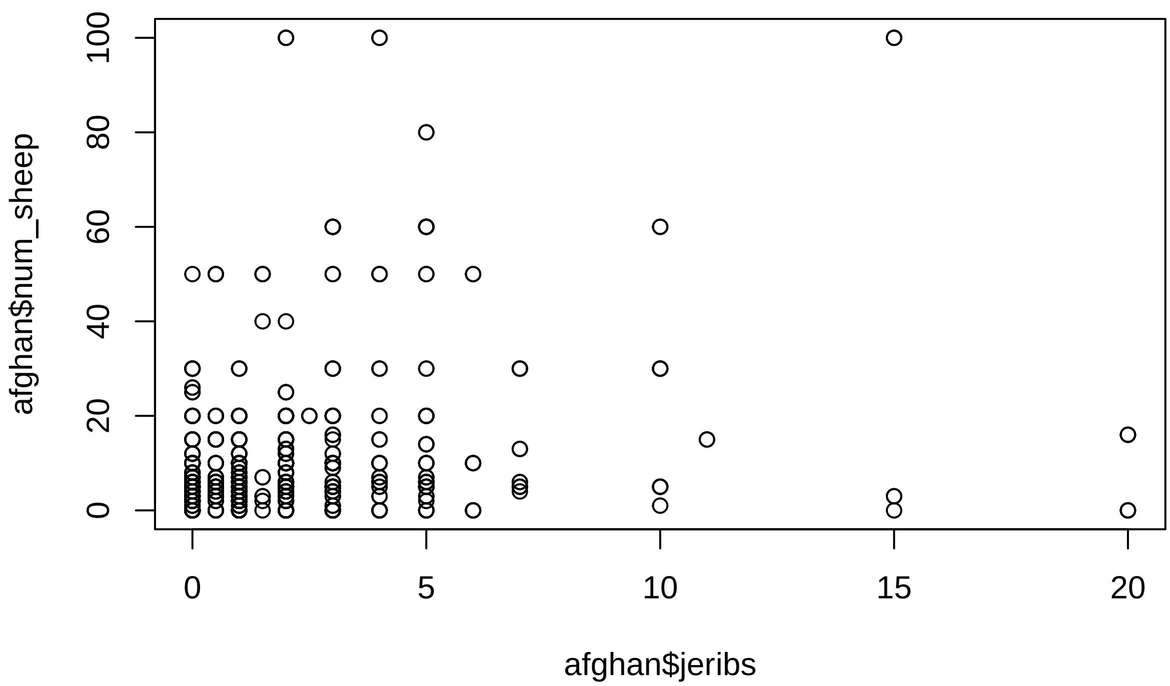
```r
kable(round(cor(afghan[, c("formal_school", "nearest_scl", "heads_child", "girl",
    "age_child", "age_head", "yrs_ed_head", "jeribs", "num_sheep", "duration_village",
    "farsi", "tajik", "farmer", "num_ppl_hh")]), digits = 2))
```

	formal_school	nearest_scl	heads_child	girl	age_child	age_head	yrs_ed_head	jeribs	nu
formal_school	1.00	-0.12	0.00	-0.10	0.16	-0.03	0.05	0.00	

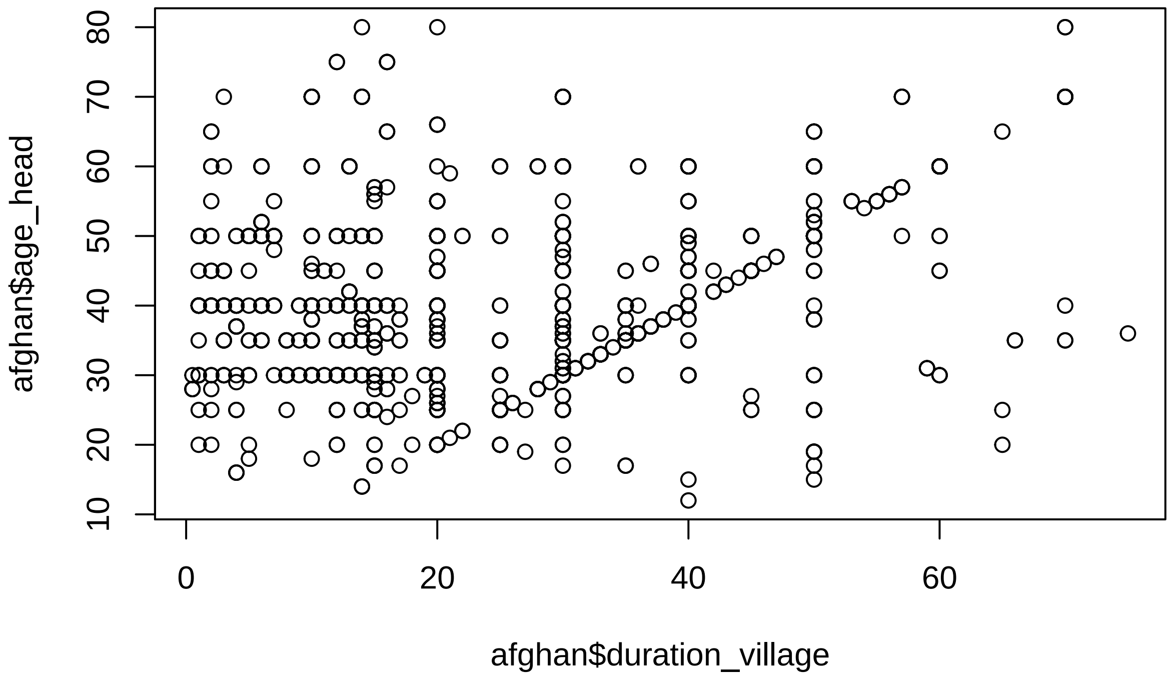
	formal_school	nearest_scl	heads_child	girl	age_child	age_head	yrs_ed_head	jeribs	nu
nearest_scl	-0.12	1.00	0.04	-0.02	0.01	0.03	-0.01	0.01	
heads_child	0.00	0.04	1.00	0.02	0.00	0.12	0.01	0.00	
girl	-0.10	-0.02	0.02	1.00	0.01	-0.01	0.00	0.03	
age_child	0.16	0.01	0.00	0.01	1.00	0.07	0.01	0.00	
age_head	-0.03	0.03	0.12	-0.01	0.07	1.00	0.02	0.08	
yrs_ed_head	0.05	-0.01	0.01	0.00	0.01	0.02	1.00	0.05	
jeribs	0.00	0.01	0.00	0.03	0.00	0.08	0.05	1.00	
num_sheep	0.13	-0.03	-0.05	0.03	0.02	0.09	0.14	0.32	
duration_village	0.02	-0.06	-0.04	0.01	0.02	0.31	0.00	0.04	
farsi	-0.02	0.02	0.02	0.03	-0.01	-0.01	-0.04	0.01	
tajik	0.03	0.06	-0.02	0.00	-0.01	0.08	0.01	-0.07	
farmer	-0.07	0.08	-0.01	-0.01	-0.03	0.13	-0.28	0.01	
num_ppl_hh	0.06	-0.08	-0.20	0.05	0.01	0.09	0.21	0.11	

none are more than 0.35. above magnitude 0.25 are: yrs head of household education and farmer -0.28, farsi and tajik -0.27, duration in village and age of household head 0.3, sheep and jerobs 0.32.

```r
plot(afghan$jeribs, afghan$num_sheep)
```



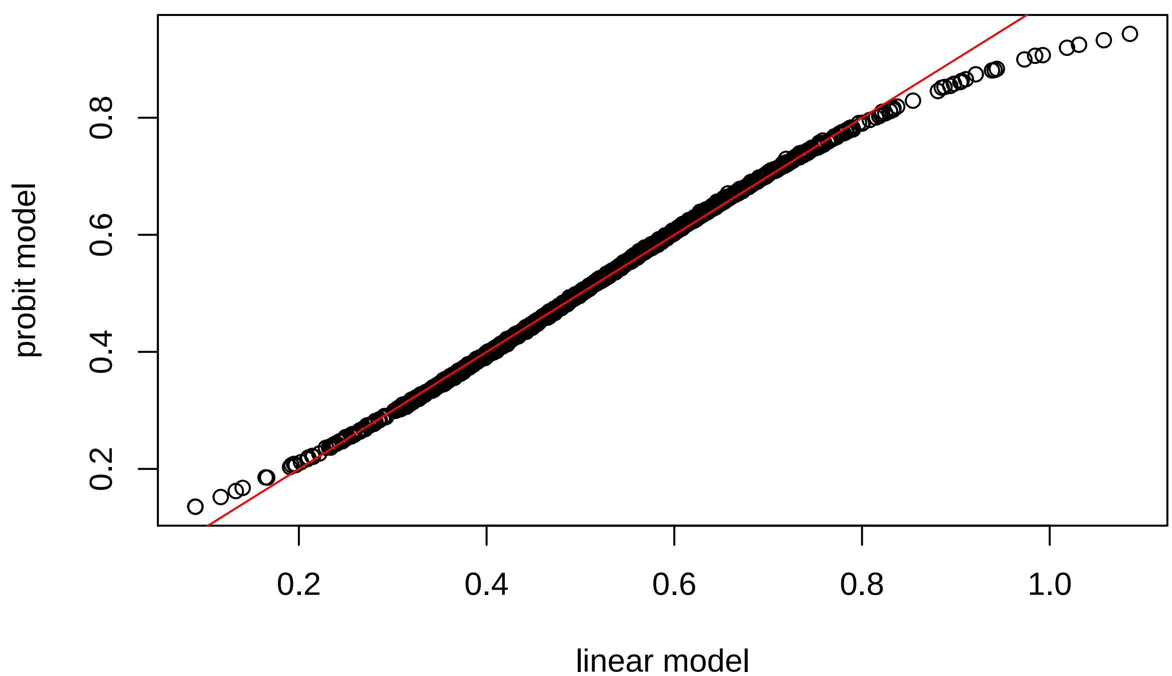

```r
plot(afghan$duration_village, afghan$age_head)
```



WHY CHOOSE PROBIT? PLOT the RESIDUALS? cross ref with balance tables fixed effects for village

nearest_scl, girl, age_child, age_head, num_sheep all significant

```
plot(predict(r2, type = "response"), predict(p2, type = "response"), xlab = "linear model",
    ylab = "probit model")
abline(a = 0, b = 1, col = "red")
```



we can see the difference in the tails

```
linearHypothesis(p2, c("tajik = 0", "farsi = 0"), test = "F")
```

```
## Linear hypothesis test
```

Table 3:

	Dependent variable:				
	formal_school				
	OLS			*probit*	
	(1)	(2)	(3)	(4)	(5)
Constant	0.671***	0.239**	0.405***	0.440***	−0.685***
	(0.036)	(0.097)	(0.102)	(0.092)	(0.258)
nearest_scl	−0.053***	−0.051***	−0.020*	−0.136***	−0.139***
	(0.011)	(0.011)	(0.011)	(0.029)	(0.029)
girl		−0.114***	−0.116***		−0.303***
		(0.025)	(0.021)		(0.065)
age_child		0.051***	0.051***		0.134***
		(0.007)	(0.006)		(0.020)
age_head		−0.003**	−0.002*		−0.008**
		(0.001)	(0.001)		(0.003)
num_sheep		0.005***	0.001		0.015***
		(0.001)	(0.001)		(0.003)
jeribs		−0.009	0.008		−0.024
		(0.006)	(0.005)		(0.016)
yrs_ed_head		0.003	0.003		0.009
		(0.003)	(0.003)		(0.009)
heads_child		0.062	−0.008		0.170
		(0.046)	(0.040)		(0.124)
duration_village		0.001	−0.001		0.002
		(0.001)	(0.001)		(0.002)
num_ppl_hh		0.006	−0.001		0.017
		(0.004)	(0.003)		(0.011)
tajik		0.042	0.037		0.112
		(0.031)	(0.027)		(0.082)
farsi		−0.014	−0.004		−0.034
		(0.032)	(0.027)		(0.084)
as.factor(clustercode)2			−0.091		
			(0.071)		
as.factor(clustercode)3			−0.615***		
			(0.072)		
as.factor(clustercode)4			0.015		
			(0.068)		
as.factor(clustercode)5			0.191***		
			(0.068)		
as.factor(clustercode)6			0.226***		
			(0.068)		

```
##
## Hypothesis:
## tajik = 0
## farsi = 0
##
## Model 1: restricted model
## Model 2: formal_school ~ nearest_scl + girl + age_child + age_head + num_sheep +
##     jeribs + yrs_ed_head + heads_child + duration_village + num_ppl_hh +
##     tajik + farsi
##
##   Res.Df Df      F Pr(>F)
## 1   1549
## 2   1547  2 1.2658 0.2823
```

```r
linearHypothesis(p2, c("num_sheep = 0", "jeribs = 0"), test = "F")
```

```
## Linear hypothesis test
##
## Hypothesis:
## num_sheep = 0
## jeribs = 0
##
## Model 1: restricted model
## Model 2: formal_school ~ nearest_scl + girl + age_child + age_head + num_sheep +
##     jeribs + yrs_ed_head + heads_child + duration_village + num_ppl_hh +
##     tajik + farsi
##
##   Res.Df Df      F    Pr(>F)
## 1   1549
## 2   1547  2 10.771 2.262e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

jeribs and sheep jointly significant farsi and tajik not.

```r
detach(afghan)
```