Appendix

Joe Backer, Elle Dodd, Dan Humphrey, Dan Peckham

3/25/2017

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

```
afghan <- read_dta("~/Documents/Stats2/pivotproject/afghandata.dta")
#rename variables
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"
))
#age by gender
afghan$age_girl <- afghan$age_child*afghan$girl
afghan$age_girl[afghan$age_girl  == 0] <- NA
afghan$age_boy <-  afghan$age_child* !afghan$girl
afghan$age_boy[afghan$age_boy  == 0] <- NA

#sheep per household member (proxy for wealth?)
afghan$sheep_per_hh_member = afghan$num_sheep / afghan$num_ppl_hh

attach(afghan)
```

```
# Create Balance Table
```

```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_ctrl <- apply(balance_variables[balance_variables$treatment == 0, ], 2, function(x) length(which(!is.n
n_trt <- apply(balance_variables[balance_variables$treatment == 1, ], 2, function(x) length(which(!is.na

# generate table
balancetable <- cbind(n_ctrl, n_trt)
# drop treatment, test score, cluster rows
balancetable <- balancetable[!rownames(balancetable) == "treatment", ]
balancetable <- balancetable[!rownames(balancetable) == "test_score_normalized",
    ]
balancetable <- balancetable[!rownames(balancetable) == "clustercode", ]


# run t.tests, skipping treatment[14]
varlist
```

```
##  [1] "heads_child"           "girl"
##  [3] "age_child"             "age_head"
##  [5] "yrs_ed_head"           "jeribs"
##  [7] "num_sheep"             "duration_village"
##  [9] "farsi"                 "tajik"
## [11] "farmer"                "num_ppl_hh"
## [13] "test_observed"         "treatment"
## [15] "clustercode"           "chagcharan"
## [17] "formal_school"         "nearest_scl"
## [19] "test_score_normalized" "age_girl"
## [21] "age_boy"               "sheep_per_hh_member"
```

```r
balance_tests <- lapply(varlist[c(1:13, 16:18, 20:22)], 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(mean_crtl = unname(x$estimate[1]), mean_trt = unname(x$estimate[2]), diff_means = unname(x$estima
        unname(x$estimate[1]), p.value = p.adjust(x$p.value, method = "bonferroni",
        n = length(x)))
}))

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)
# Show table
kable(balancetable)
```

	n_ctrl	n_trt	mean_crtl	mean_trt	diff_means	p.value

	n_ctrl	n_trt	mean_crtl	mean_trt	diff_means	p.value
heads_child	730	830	0.911	0.927	0.016	1.000
girl	730	830	0.456	0.478	0.022	1.000
age_child	730	830	8.321	8.322	0.001	1.000
age_head	730	830	40.219	40.090	-0.129	1.000
yrs_ed_head	730	830	3.101	3.531	0.431	0.165
jeribs	730	830	1.510	1.498	-0.011	1.000
num_sheep	730	830	6.404	9.586	3.181	0.000
duration_village	730	830	27.662	30.172	2.509	0.014
farsi	730	830	0.205	0.210	0.004	1.000
tajik	730	830	0.204	0.239	0.034	0.914
farmer	730	830	0.729	0.707	-0.022	1.000
num_ppl_hh	730	830	7.905	8.741	0.835	0.000
test_observed	730	830	0.925	0.925	0.001	1.000
chagcharan	730	830	0.429	0.663	0.234	0.000
formal_school	730	830	0.264	0.731	0.467	0.000
nearest_scl	730	830	3.149	2.881	-0.268	0.000
age_girl	333	397	8.327	8.332	0.005	1.000
age_boy	397	433	8.315	8.312	-0.003	1.000
sheep_per_hh_member	730	830	0.817	1.142	0.325	0.000

```r
# Attrition summary stats
sum_list <- list(afghan$test_observed, afghan$test_observed[treatment == 1],
    afghan$test_observed[treatment == 0])
n <- sapply(sum_list, function(x) length(which(!is.na(x))))
mean <- sapply(sum_list, mean, na.rm = T)
sd <- sapply(sum_list, sd, na.rm = T)
sum_table <- cbind(n, mean, sd)
sum_table <- round(sum_table, digits = 3)
rownames(sum_table) <- c("% Test Taken - All", "% Test Taken - Treatment", "% Test Taken - Control")
kable(sum_table)
```

	n	mean	sd
% Test Taken - All	1560	0.925	0.263
% Test Taken - Treatment	830	0.925	0.263
% Test Taken - Control	730	0.925	0.264

```r
# creating a dataset of only the attritted
afghanattrition <- afghan[!complete.cases(afghan), ]

# comparisons of treatment and control for attritted only

# omits test_observed[13], treatment[14], clustercode[15], and
# test_score[19]
attrition_by_treatment <- lapply(varlist[c(1:12, 16:18, 20:22)], function(x) {
    t.test(as.formula(paste(x, "treatment", sep = "~")), data = afghanattrition,
        alternative = "two.sided", mu = 0, paired = FALSE, var.equal = FALSE,
        conf.level = 0.95)
```

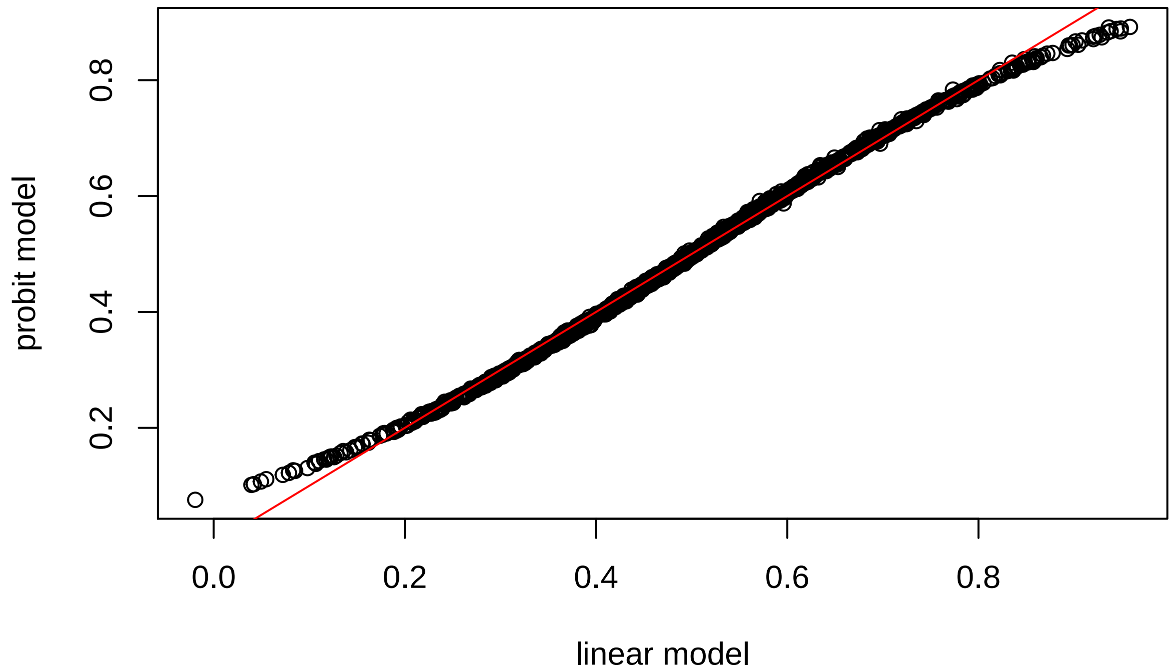
```
})

# create table
attrition_table <- t(sapply(attrition_by_treatment, function(x) {
    c(x$data.name, mean_crtl = unname(x$estimate[1]), mean_trt = unname(x$estimate[2]),
        diff_means = unname(x$estimate[2]) - unname(x$estimate[1]), p.value = p.adjust(x$p.value,
            method = "bonferroni", n = length(x)))
}))
rownames(attrition_table) <- attrition_table[, 1]
attrition_table <- attrition_table[, -c(1)]
attrition_table <- data.frame(attrition_table, stringsAsFactors = FALSE)
attrition_table[] <- lapply(attrition_table, function(x) as.numeric(as.character(x)))
rownames(attrition_table) <- varlist[c(1:12, 16:18, 20:22)]
kable(attrition_table)
```

	mean_crtl	mean_trt	diff_means	p.value
heads_child	0.9109589	0.9265060	0.0155471	1.0000000
girl	0.4561644	0.4783133	0.0221489	1.0000000
age_child	8.3205479	8.3216867	0.0011388	1.0000000
age_head	40.2191781	40.0903614	-0.1288166	1.0000000
yrs_ed_head	3.1006849	3.5313253	0.4306404	0.1648415
jeribs	1.5095890	1.4981928	-0.0113963	1.0000000
num_sheep	6.4041096	9.5855422	3.1814326	0.0000010
duration_village	27.6623288	30.1716867	2.5093580	0.0137177
farsi	0.2054795	0.2096386	0.0041591	1.0000000
tajik	0.2041096	0.2385542	0.0344446	0.9137745
farmer	0.7287671	0.7072289	-0.0215382	1.0000000
num_ppl_hh	7.9054795	8.7409639	0.8354844	0.0000056
chagcharan	0.4287671	0.6626506	0.2338835	0.0000000
formal_school	0.2643836	0.7313253	0.4669417	0.0000000
nearest_scl	3.1492138	2.8811876	-0.2680262	0.0000229
age_girl	8.3273273	8.3324937	0.0051664	1.0000000
age_boy	8.3148615	8.3117783	-0.0030832	1.0000000
sheep_per_hh_member	0.8172399	1.1422055	0.3249656	0.0000787

No significant correlation between demographic variables greater than 0.35.

WHY CHOOSE PROBIT? PLOT the RESIDUALS? cross ref with balance tables

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




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

```
## Linear hypothesis test
##
## 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 + farmer + sheep_per_hh_member + chagcharan
##
##   Res.Df Df      F Pr(>F)
## 1   1546
## 2   1544  2 2.2791 0.1027
```

```
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 +
```

Table 4: Linear OLS model

	Dependent variable:	
	formal_school	
	(1)	(2)
Constant	0.231**	0.436***
	(0.100)	(0.111)
nearest_scl	−0.052***	−0.018*
	(0.010)	(0.011)
girl	−0.112***	−0.116***
	(0.024)	(0.021)
age_child	0.051***	0.051***
	(0.007)	(0.007)
age_head	−0.003**	−0.001
	(0.001)	(0.001)
num_sheep	0.004	0.001
	(0.004)	(0.004)
jeribs	−0.008	0.008*
	(0.006)	(0.005)
yrs_ed_head	0.005	0.002
	(0.004)	(0.003)
heads_child	0.038	−0.010
	(0.047)	(0.039)
duration_village	0.0004	−0.001
	(0.001)	(0.001)
num_ppl_hh	0.002	−0.001
	(0.005)	(0.004)
tajik	0.059*	0.042
	(0.030)	(0.027)
farsi	−0.012	−0.0001
	(0.031)	(0.027)
farmer	−0.057**	−0.043*
	(0.029)	(0.025)
sheep_per_hh_member	−0.011	−0.004
	(0.028)	(0.028)
chagcharan	0.214***	
	(0.025)	
clustercode fixed effects?	No	Yes
Observations	1,560	1,560
R^2	0.120	0.339
Adjusted R^2	0.111	0.329

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5: Probit model

	Dependent variable:	
	formal_school	
	(1)	(2)
Constant	-0.737^{***}	-0.331
	(0.276)	(0.363)
nearest_scl	-0.146^{***}	-0.070^{*}
	(0.030)	(0.038)
girl	-0.309^{***}	-0.411^{***}
	(0.067)	(0.075)
age_child	0.138^{***}	0.174^{***}
	(0.020)	(0.023)
age_head	-0.008^{**}	-0.006
	(0.003)	(0.004)
num_sheep	0.011	0.0004
	(0.010)	(0.011)
jeribs	-0.022	0.037^{*}
	(0.016)	(0.019)
yrs_ed_head	0.013	0.008
	(0.010)	(0.011)
heads_child	0.114	-0.054
	(0.127)	(0.143)
duration_village	0.001	-0.003
	(0.002)	(0.003)
num_ppl_hh	0.008	-0.002
	(0.013)	(0.015)
tajik	0.160^{*}	0.158^{*}
	(0.084)	(0.095)
farsi	-0.032	-0.00003
	(0.085)	(0.098)
farmer	-0.154^{**}	-0.156^{*}
	(0.078)	(0.088)
sheep_per_hh_member	-0.022	0.019
	(0.082)	(0.089)
chagcharan	0.576^{***}	
	(0.070)	
clustercode fixed effects?	No	Yes
Observations	1,560	1,560
Note:	*p<0.1; **p<0.05; ***p<0.01	

```
##     jeribs + yrs_ed_head + heads_child + duration_village + num_ppl_hh +
##     tajik + farsi + farmer + sheep_per_hh_member + chagcharan
##
##   Res.Df Df      F Pr(>F)
## 1   1546
## 2   1544  2 1.4186 0.2424
```