Wealthy Women Can Afford to Have More Children. Do They?

Grant Donovan, Manny Brinson

University of Central Florida, College of Business

ECO4422: Econometrics II

Dr. Richard Hofler

April 11th, 2023

A. Introduction

The subject of childbirth has quickly become one of the most pressing topics facing economists and policy analysts in our country today. The United States birth rate is currently the lowest it's ever been, and experts are warning this may lead to worker shortages and dwindling tax bases in the future. Most of the research behind this trend has focused on its relationship to the recent demographic changes of American mothers. But in this study, we wanted to focus on a different set of factors: economic factors. Having a child is undoubtedly a large investment; an investment economically advantaged women are in a much better position to make. So do they? In this report, we seek to answer that question and many like it. It is our hope that these results shed a greater light on the microeconomic decision-making behind having a child, which can then be applied by leaders seeking to better understand and manipulate the national birth rate.

B. Data

Response Variable:

- Number of Own Children in Household (nchild) – Each individual's total number of their own children residing in their household at the time of the survey.

Independent Variables:

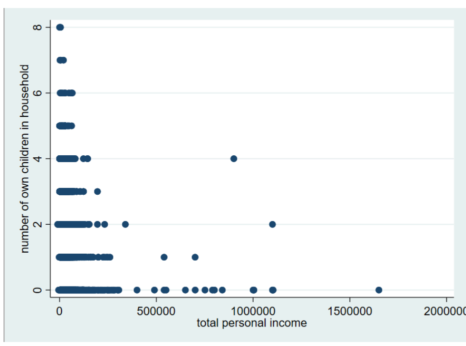
- Total Personal Income (inctot) – Each individual's total pre-tax personal income from all sources the previous calendar year.
- Marital Status (marst_dummy) – A binary variable equal to 1 if the individual is married and 0 if not.
- Usual Hours Worked / Week (uhrsworkt) – Each individual's usual hours worked per week at all jobs.
- Health (health) – Each individual rated the quality of their health on a scale from 1 being poor to 5 being excellent.

- Additional Household Income (ahhincome) – The total income of each individual's household minus their own personal income.
- Educational Attainment (educ) – Each individual's highest level of educational attainment as ranked on a scale from 1 being grades 1-4 to 12 being a doctorate degree.

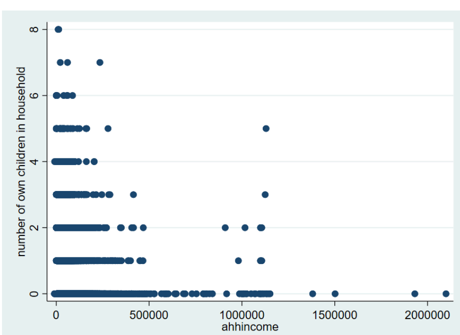
<u>Data Source:</u>

- Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. *Integrated Public Use Microdata Series, Current Population Survey: Version 10.0*. Minneapolis, MN: IPUMS, 2022. https://doi.org/10.18128/D030.V10.0

C. **<u>Plots of Dependent Variable and Independent Variables</u>**



- Response plotted against X_1 (Total Personal Income). Potentially displays a non-linear relationship.
- The natural log of this variable was used in the final regression based on the comparison of the BIC scores between models in figure 1. (See Appendix)



- Response plotted against X_5 (Additional Household Income). Potentially displays a non-linear relationship.
- The natural log of this variable was used in the final regression based on the comparison of the BIC scores between models in figure 1. (See Appendix)

- The graph relating the response variable to X_3 (Usual Hours Worked / Week) does not display any non-linear relationship. (See Appendix)
- See appendix for graphs relating the response to the categorical means of binary variable X_2 (Marital Status) and ordinal variables X_4 (Health) and X_6 (Educational Attainment).

D. Model Specification and Regression Estimation Results

Negative binomial regression — Number of obs = 21,049

Dispersion: mean — Prob > chi2 = 0.0000

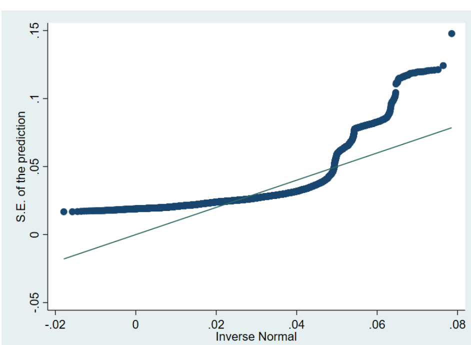
Log likelihood = -18346.791 — Pseudo R2 = 0.0960

nchild	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
log_inctot	-.0018868	.0061971	-0.30	0.761	-.0140329	.0102593
marst_dummy	1.416815	.0259037	54.70	0.000	1.366045	1.467585
uhrsworkt	-.0030196	.0010984	-2.75	0.006	-.0051725	-.0008666
log_health	-.332667	.0460571	-7.22	0.000	-.4229373	-.2423968
log_ahhincome	-.0953746	.0028067	-33.98	0.000	-.1008756	-.0898736
log_educ	-1.427391	.0549236	-25.99	0.000	-1.535039	-1.319742
_cons	3.184396	.1284288	24.80	0.000	2.932681	3.436112
/lnalpha	-.4961804	.0488092			-.5918447	-.4005161
alpha	.6088518	.0297176			.5533057	.6699742

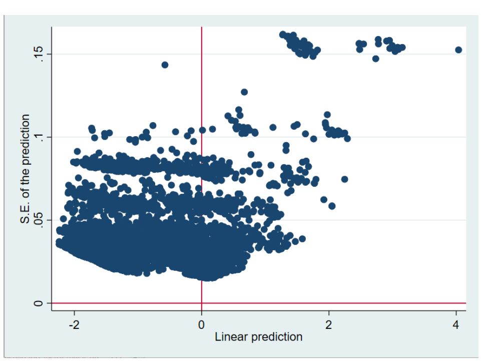
E. Checks of the Adequacy of the Model

```
. gen leverage = yhat / (1-yhat) * model_residual^2
. drop if leverage > 0.1
(4 observations deleted)
. drop if leverage < -0.1
(2 observations deleted)
```

- Six high leverage observations were removed. Five had exceedingly large incomes. The sixth had an unrealistic Usual Hours Worked / Week.



- The model's residuals do not lie along the normal quantiles plot line of equality. This indicates that the error term is not normally distributed, and significant variables have been omitted from our model.



- The model's residuals all lie above 0, indicating that it consistently underestimates the correct values of the response. The Negative Binomial Regression Model requires neither homoscedasticity nor that e ~ N.

F. Inference and Interpretations

- Marital Status (**β_2 = 1.416085**) – The ratio of the response variable between women who are married and women who are not is = exp(1.417073) = 4.121. Married women are inferred to be 4.121x more likely to have children than unmarried women.
- Usual Hours Worked / Week (**β_3 = -0.0030196**) **–** For every 1% increase in a woman's usual hours worked / week, she is 0.003% less likely to have an additional child in her household.
- Health Status (**β_4 = -0.332667**) **–** For every 1% increase in a woman's health status, she is 0.333% less likely to have an additional child in her household.
- Additional Household Income (**β_5 = -0.0953746**) **–** For every 1% increase in a woman's additional household income, she is 0.095% less likely to have an additional child in her household.
- Educational Attainment (**β_6 = -1.427391**) **–** For every 1% increase in a woman's educational attainment, she is 1.43% less likely to have an additional child in her household.

G. Summary

In this project, we sought to quantify the effects that economic factors have on an individual's decision to have children. To control for demographic factors, we restricted our sample to women aged 20-29. Our response was count data with a variance greater than its mean . Therefore, we modeled it as a Negative Binomial Distribution – for which regression requires the use of a log-linear link function. We then took the natural logs of our regressor variables which were either ordinal or displayed non-linear relationships with the response. Likelihood Ratio tests indicated that our unrestricted model including all six variables was preferred. No multicollinearity was apparent in any VIF statistics. (See Appendix)

After estimation, we were surprised to see total personal income with an insignificant coefficient. We conclude that it is not income which reduces a woman's decision to have children, but rather her innate workforce capacity and ambition – of which educational attainment is a signifier. It is also worth noting that the checks of model adequacy all suggest this model omits important variables. This is supported by the small pseudo-R^2 values of all the models we estimated. Economic factors must not influence women's decision to have children as much as other kinds, such as socio-cultural factors, do.

H. Appendix

Figure 1 Log-Linear Regression Model vs. Log-Log Regression Model

Negative binomial regression — Number of obs = **21,049**; LR chi2(**6**) = **3780.65**; Prob > chi2 = **0.0000**; Pseudo R2 = **0.0931**

Dispersion: **mean**
Log likelihood = **-18404.082**

nchild	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
inctot	-2.57e-07	3.88e-07	-0.66	0.508	-1.02e-06	5.04e-07
marst_dummy	1.194738	.0238469	50.10	0.000	1.147999	1.241478
uhrswork	-.0002199	.0010846	-0.20	0.839	-.0023456	.0019059
health	-.0764029	.0131188	-5.82	0.000	-.1021152	-.0506906
ahhincome	-5.66e-06	2.62e-07	-21.57	0.000	-6.17e-06	-5.15e-06
educ_recode	-.2453923	.0078235	-31.37	0.000	-.260726	-.2300586
_cons	1.513352	.0833924	18.15	0.000	1.349906	1.676798
ln(exposu~r)	1	(exposure)				
/lnalpha	-.5661685	.0518886			-.6678683	-.4644687
alpha	.5676964	.029457			.5128006	.6284689

LR test of alpha=0: chibar2(01) = **696.50** Prob >= chibar2 = **0.000**

. **estat ic**

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	21,049	-20294.41	-18404.08	8	36824.16	36887.8

Negative binomial regression — Number of obs = **21,049**; Wald chi2(**6**) = **3927.16**; Prob > chi2 = **0.0000**; Pseudo R2 = **0.0961**

Dispersion: **mean**
Log pseudolikelihood = **-18344.7**

nchild	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
log_inctot	-.0023908	.0058721	-0.41	0.684	-.0138999	.0091183
marst_dummy	1.417073	.0251504	56.34	0.000	1.367779	1.466367
uhrswork	-.0030882	.0010735	-2.88	0.004	-.0051923	-.0009842
log_health	.1875277	.0252518	7.43	0.000	.1380351	.2370202
log_ahhincome	-.0954642	.0028446	-33.56	0.000	-.1010395	-.0898888
log_educ	-1.423467	.072665	-19.59	0.000	-1.565888	-1.281046
_cons	2.623271	.1508198	17.39	0.000	2.32767	2.918873
ln(exposur~r)	1	(exposure)				
/lnalpha	-.4972074	.0525071			-.6001194	-.3942953
alpha	.6082268	.0319362			.5487461	.6741549

.
. **estat ic**

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	21,049	-20294.41	-18344.7	8	36705.4	36769.04

Figure 2 Response Plotted Against the Category Means of X_2 (Marital Status)

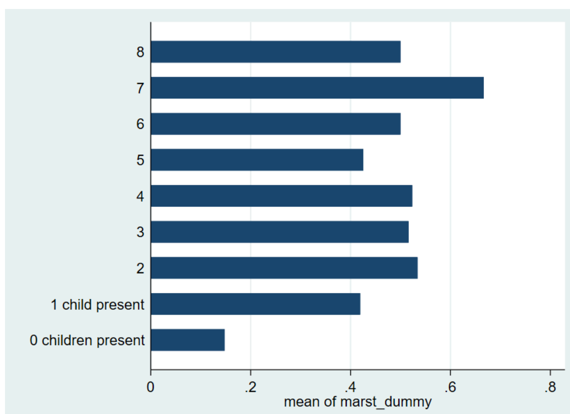


Figure 3 Response Plotted Against X_3 (Usual Hours Worked / Week)

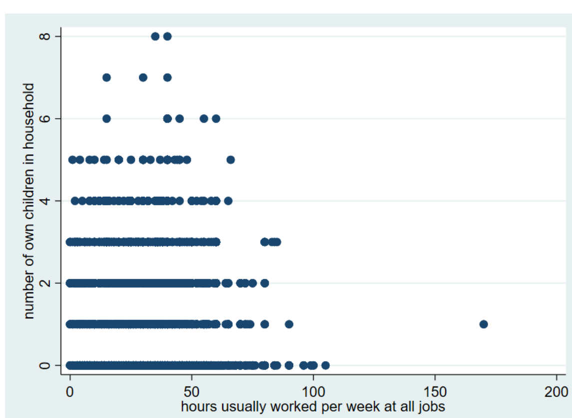


Figure 4 **Response Plotted Against the Category Means of X_4 (Health)**

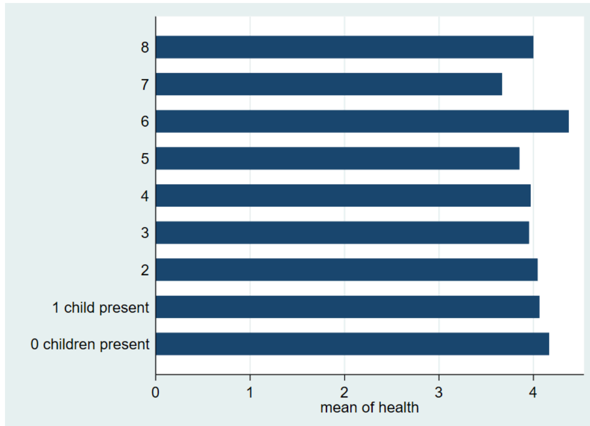


Figure 5 **Response Plotted Against the Categorical Means of X_6 (Educational Attainment)**

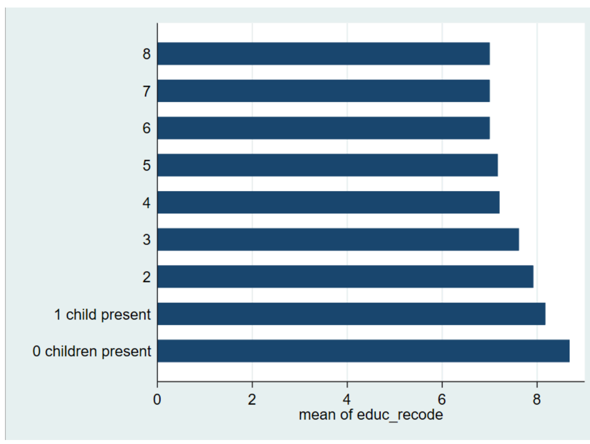


Figure 6 **Poisson Regression Model vs. Negative Binomial Regression Model**

```
Poisson regression                                 Number of obs =    21,049
                                                   Wald chi2(6)  =   3335.85
                                                   Prob > chi2   =    0.0000
Log pseudolikelihood = -18753.574                  Pseudo R2     =    0.1280
```

nchild	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
log_inctot	-.0119592	.0055449	-2.16	0.031	-.022827	-.0010913
marst_dummy	1.342672	.025137	53.41	0.000	1.293404	1.39194
uhrswork t	-.0044749	.0010379	-4.31	0.000	-.0065091	-.0024408
log_health	.1961994	.0245394	8.00	0.000	.1481032	.2442957
log_ahhincome	-.0908412	.0027721	-32.77	0.000	-.0962744	-.085408
log_educ	-.9163618	.049041	-18.69	0.000	-1.01248	-.8202432
_cons	1.693703	.1068296	15.85	0.000	1.484321	1.903086
ln(exposur~r)	1	(exposure)				

. estat ic

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	21,049	-21506.92	-18753.57	7	37521.15	37576.83

```
Negative binomial regression                       Number of obs =    21,049
                                                   Wald chi2(6)  =   3927.16
Dispersion: mean                                   Prob > chi2   =    0.0000
Log pseudolikelihood = -18344.7                    Pseudo R2     =    0.0961
```

nchild	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
log_inctot	-.0023908	.0058721	-0.41	0.684	-.0138999	.0091183
marst_dummy	1.417073	.0251504	56.34	0.000	1.367779	1.466367
uhrsworkt	-.0030882	.0010735	-2.88	0.004	-.0051923	-.0009842
log_health	.1875277	.0252518	7.43	0.000	.1380351	.2370202
log_ahhincome	-.0954642	.0028446	-33.56	0.000	-.1010395	-.0898888
log_educ	-1.423467	.072665	-19.59	0.000	-1.565888	-1.281046
_cons	2.623271	.1508198	17.39	0.000	2.32767	2.918873
ln(exposur~r)	1	(exposure)				
/lnalpha	-.4972074	.0525071			-.6001194	-.3942953
alpha	.6082268	.0319362			.5487461	.6741549

.

. estat ic

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	21,049	-20294.41	-18344.7	8	36705.4	36769.04

Figure 7 Likelihood-Ratio Test Restricted vs. Saturated Model

```
. lrtest restricted_model saturated_model

Likelihood-ratio test
Assumption: restricted_m~l nested within saturated_mo~l

 LR chi2(2) = 914.65
Prob > chi2 = 0.0000
```

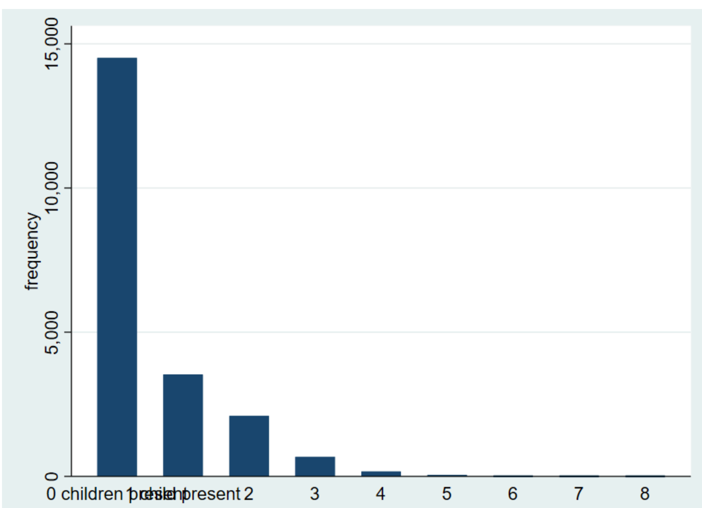
- See code below for full model specifications

Figure 8 VIF Collinearity Diagnostics

Variable	VIF	SQRT VIF	Tolerance	R-Squared
nchild	1.21	1.10	0.8243	0.1757
inctot	1.12	1.06	0.8895	0.1105
marst_dummy	1.15	1.07	0.8721	0.1279
uhrsworkt	1.08	1.04	0.9277	0.0723
health	1.02	1.01	0.9775	0.0225
ahhincome	1.04	1.02	0.9659	0.0341
educ_recode	1.18	1.09	0.8482	0.1518
Mean VIF	1.11			

Variable	VIF	SQRT VIF	Tolerance	R-Squared
nchild	1.23	1.11	0.8115	0.1885
log_inctot	1.12	1.06	0.8921	0.1079
marst_dummy	1.22	1.10	0.8211	0.1789
uhrsworkt	1.08	1.04	0.9224	0.0776
log_health	1.02	1.01	0.9833	0.0167
log_ahhincome	1.11	1.05	0.9043	0.0957
log_educ	1.13	1.06	0.8865	0.1135
Mean VIF	1.13			

Figure 9 Distribution of the Response Variable & Summary Statistics




```
                number of own children in household
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs              21,049
25%            0              0       Sum of wgt.      21,049

50%            0                      Mean           .5109981
                        Largest       Std. dev.      .8978548
75%            1              7
90%            2              7       Variance       .8061433
95%            2              8       Skewness       1.987506
99%            4              8       Kurtosis       7.354516
```

Figure 10 STATA Code

```
. cd "C:\Users\14029\Dropbox\My PC (DESKTOP-R1GC0QH)\Downloads\Econometrics Research Project"
. use "C:\Users\14029\Dropbox\My PC (DESKTOP-R1GC0QH)\Downloads\Econometrics Research
Project\cps_00010.dta"
. browse
. keep if sex == 2 & age >= 20 & age <= 29 & inctot != 999999999 & inctot != 999999998 & nchild
<= 8 >& uhrsworkt != 997 & uhrsworkt != 999 & marst != 9 & popstat !=2
```

```
. generate ahhincome = hhincome - inctot
. recode marst (2 = 0) (3 = 0) (4 = 0) (5 = 0) (6 = 0) (7 = 0), generate(marst_dummy)
. recode educ (002 = 0) (010 = 1) (020 = 2) (030 = 3) (040 = 4) (050 = 5) (060 = 6) (071 = 6)
(073 = >7) (081 = 8) (091 = 9) (092 = 9) (111 = 10) (123 = 11) (124 = 11) (125 = 12),
generate(educ_recode)
. recode health (1 = 5) (2 = 4) (4 = 2) (5 = 1)
. generate exposure_var = 1
. graph bar (count), over(nchild)
. summarize nchild, detail
. sort nchild
. ssc install egenmore
. by nchild: summarize inctot
. graph bar (mean) inctot, over(nchild) horizontal
. twoway scatter nchild inctot
. by nchild: summarize marst_dummy
. graph bar (mean) marst_dummy, over(nchild) horizontal yscale(reverse)
. by nchild: summarize uhrsworkt
. graph bar (mean) uhrsworkt, over(nchild) horizontal
. twoway scatter nchild uhrsworkt
. by nchild: summarize health
. graph bar (mean) health, over(nchild) horizontal yscale(reverse)
. by nchild: summarize ahhincome
. graph bar (mean) ahhincome, over(nchild) horizontal
. twoway scatter nchild ahhincome
. by nchild: summarize(educ_recode)
. graph bar (mean) educ_recode, over(nchild) horizontal yscale(reverse)
. generate log_inctot = log(inctot)
. recode log_inctot (. = 0)
. generate log_health = log(health)
. generate log_ahhincome = log(ahhincome)
. recode log_ahhincome (. = 0)
. generate log_educ = log(educ_recode)
. recode log_educ (. = 0)
. nbreg nchild inctot marst_dummy uhrsworkt ahhincome, exposure(exposure_var)
. estat ic
. estimates store basic_nbreg_model
. gnbreg nchild inctot marst_dummy uhrsworkt ahhincome, exposure(exposure_var)
. estat ic
. poisson nchild inctot marst_dummy uhrsworkt ahhincome, exposure(exposure_var)
. estat ic
. nbreg nchild inctot marst_dummy uhrsworkt health ahhincome educ_recode, exposure(exposure_var)
. estat ic
. estimates store sat_nbreg_model
. lrtest sat_nbreg_model basic_nbreg_model
. corr inctot marst_dummy uhrsworkt ahhincome educ_recode
. collin inctot marst_dummy uhrsworkt ahhincome educ_recode, corr
. nbreg nchild log_inctot marst_dummy uhrsworkt log_ahhincome, exposure(exposure_var)
. estat ic
. estimates store comp_nbreg_model
. nbreg nchild log_inctot marst_dummy uhrsworkt log_health log_ahhincome log_educ,
>exposure(exposure_var)
. estat ic
. estimates store comp_sat_nbreg_model
. lrtest comp_sat_nbreg_model comp_nbreg_model
. corr log_inctot marst_dummy uhrsworkt log_health log_ahhincome log_educ
. collin log_inctot marst_dummy uhrsworkt log_health log_ahhincome log_educ, corr
. predict model_residuals, stdp
. qnorm model_residuals
. predict yhat, xb
. twoway scatter model_residuals yhat, xline(0) yline(0)
. gen leverage = yhat / (1-yhat) * model_residual^2
. drop if leverage > 0.1
. drop if leverage < -0.1
. nbreg nchild log_inctot marst_dummy uhrsworkt log_health log_ahhincome log_educ,
>exposure(exposure_var) vce(robust)
. predict model_residuals2, stdp
. qnorm model_residuals2
. nbreg nchild age race log_inctot marst_dummy uhrsworkt log_health log_ahhincome log_educ,
>exposure(exposure_var) vce(robust)
. estat ic
. predict yhat2, xb
. twoway scatter model_residuals2 yhat2, xline(0) yline(0)
```