Econometrics II: Data Management and Handling Project

By Toshan Majumdar (ID:0086133)

Date: 6 February, 2022

1) Dataset:

We have used The LISS panel (Longitudinal Internet studies for the Social Sciences) dataset for our project, which consists of survey questionnaires from 7500 individuals over 5000 households in the Netherlands. The panel data stored by CentERdata requires permission for data handling. The longitudinal study is conducted every year, with the aim of following the socio-economic conditions of its panel members, who are compensated monetarily for their participation.

Our questionnaires of interest among the core studies are: Work and Schooling, Family and Household, Economic Situation: Income, and Region and Ethnicity. Our econometric modelling is based on the raw data of the longitudinal waves from 2008-2020, which is downloaded in .dta format, along with their codebooks.

2) Data Handling:

We have used the Work and Schooling questionnaire to obtain data on educational levels, and employment status.

2.1) Educational Levels:

We have considered the participant responses to question ***005 (*006 for 2019)** as their highest level of education received in the Netherlands. This variable has been converted into a categorical variable based on their responses: 0: No Education (values 1,2), 1: Completed School (values 3-15), 2: Completed University (values 16-21), 3: Completed Advanced Degree (values 22-26), 4:Other (values 27).

Missing Data: We handled the large proportion of non-responses for the year 2019 (96.3%) by replacing missing values with the education level of the previous year.

2.2) Employment Status:

We have considered the participant responses to the question ***001** (does respondent have paid work: yes/no) as their employment status. Missing data is not an issue since the average non response rate is: 0.17% .

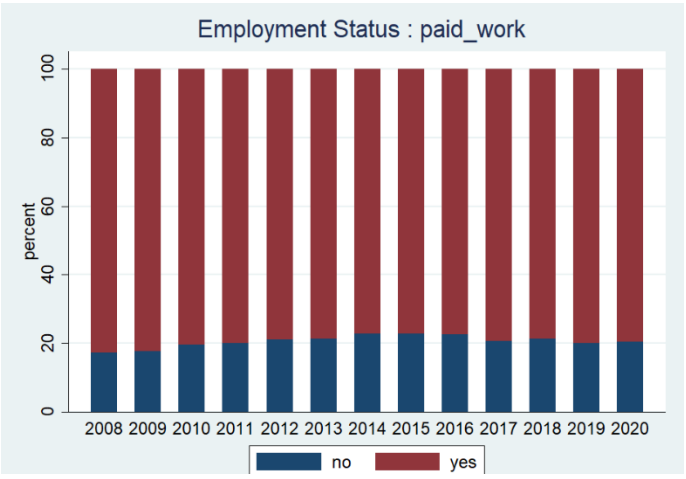

FIGURE 1: YEARWISE DISTRIBUTION OF PAID WORK STATUS

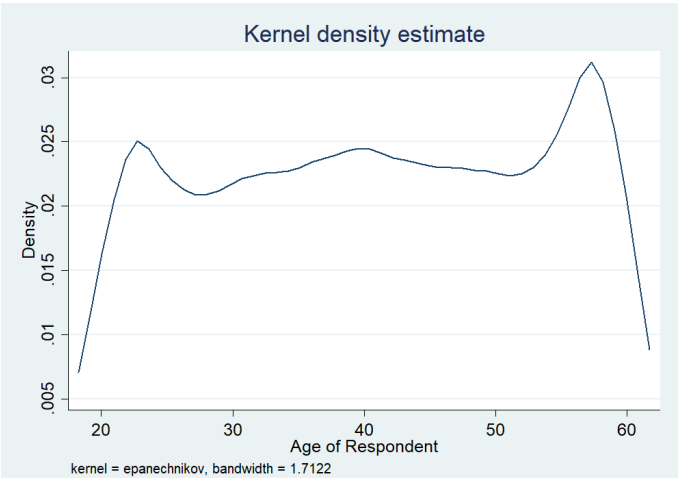

FIGURE 2: AGE DISTRIBUTION OF RESPONDENTS

2.3) Participant Year of Birth:

We have considered the participant responses to the question ***002** as their year of birth. Missing data is not an issue since the average non response rate is: 0% . The proportion of outliers (birth_year == 1900,1913) is insignificant. We will later compute the age of the participant by subtracting their birth year from the current survey wave.

We have used the **Religion and Ethnicity questionnaire** (78305 responses | 15653 unique individuals) to obtain data on nationality and views on abortion.

2.4) Nationality:

We have considered the responses to question ***079** (Which languages did you speak growing up?) as an indicator of Dutch nationality rather than question ***043** (What is your nationality?), since the latter has been omitted from surveys after 2010.

Missing Data: We handled the large proportion of non-responses for the year 2014 by replacing missing values with the nationality of the previous year since it is a time invariant variable.

2.5) Abortion:

We consider responses to the question ***105** (Do you consider it good that abortion is permitted? Yes/No) as an indicator on the individual's attitudes on large families.

Missing Data: Since this question has been omitted from surveys after 2018, we have replaced the missing responses with the reply from previous years, with the assumption that such a belief would not change in the years 2019,2020.

We have used the **Family and Household questionnaire** (78264 responses | 15608 unique individuals) to obtain data on gender, partner characteristics and family size.

2.6) Gender:

We consider responses to the question ***003** (Respondent Gender) as a self-reported indicator of the individual's gender. Missing data is not an issue since the average non response rate is: 0.07% .

2.7) Partner Characteristics**:** We have considered responses to questions ***024**, ***026**, and ***032** as self-reported indicators to the individual's partner, partner's birth year, and partner's gender. Since the non-response rate of these survey questions is high , we have decide to exclude such variables from our analysis.

2.9) Family Size:

In order to calculate family size, responses to the question ***036** (How many children do you have? 1-15) was used. High non response rate was observed for the years 2009-2014, after which the survey question was modified by the researchers into: ***455** (How many living children do you have ?)

Missing Values: In order to deal with the missing values from 2009-2014, we calculated the no. of children based on the responses given regarding the participant's children's gender (***068-*082**). We also tried to include childcare allowance status (***385**), but due to high non response rate (89.6%), the external validity of the analysis is not strong.

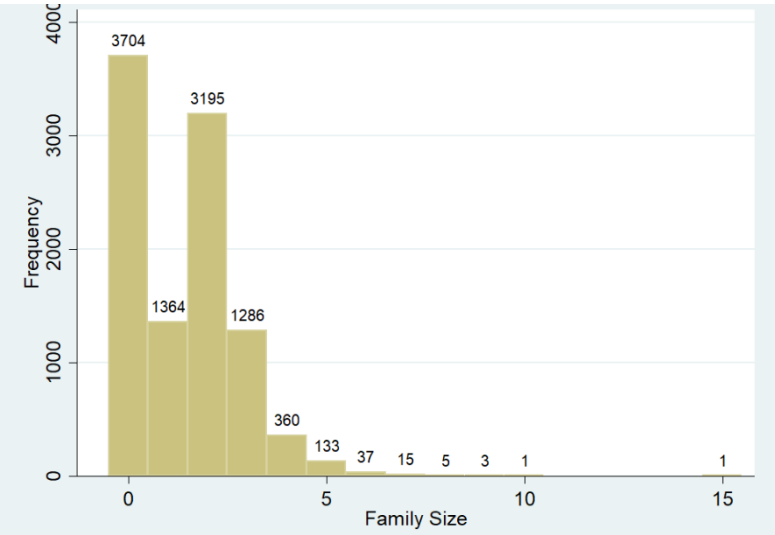

FIGURE 1: DISTRIBUTION OF FAMILY SIZE OF RESPONDENTS. DATA HAS A POSITIVE SKEW DUE TO MAJORITY OF FAMILIES HAVING 1-3 CHILDREN

We have used the **Economic Situation: Income** (72,012 Reponses | 14264 individuals) to obtain data on financial situation and retirement status.

2.10) Financial Situation:

We have considered responses to the question ***252** (How would you describe the financial situation of your household at this moment?) as an indicator of the household family situation. We have converted participant responses to a categorical variable consisting of: 1: Good Financial Situation (values 4-5), 2: Bad Financial Situation (values 1-3). Missing data is not an issue since the average non response rate is 12.58%.

2.11) Retirement Status:

We have considered responses to the question ***065** (Were you on early retirement in YEAR or for a part of YEAR?) as an indicator of retirement status of the participant. Missing data is an issue since the average non response rate is 45.8%. We have used age as a proxy for retirement status (age<60) in the final analysis due to the missing data.

3) Merging Data:

In the first stage, we appended the various longitudinal waves (2008-2020) for each distinct survey, after creating a variable indicating year (obtained through the .dta file name). This resulted in 4 datasets:
ws_combine.dta (77,227 observations | 15148 unique respondents) (Table 1.4)
family_combine.dta (78264 responses | 15608 unique individuals) (Table 1.1)
ethnic_combine.dta (78305 responses | 15653 unique individuals) (Table 1.2)
income_combine.dta(72,012 responses | 14264 unique individuals) (Table 1.3)

The next stage, is to merge the 4 datasets into a single .dta file:combine_dta, using the combination of survey_year and respondent ID**: noemem_encr** (unique to each participant, assigned randomly by the researchers)as a primary key.

The final stage, after creating a single merged dataset: **combine.dta** (Table 2.1) was imputing the missing values of time invariant characteristics of the sample from the non-missing values within each participant's observations. This was done for the variables: **birth_year, nationality,** and **abortion**. We also removed individual observations whose age is less than 20 and more than 60, to focus on those individuals with a tangible relation between employment and family size. The resultant panel data contains 48,382 observations of 11167 unique participants and 16 participant characteristics (variables).

The resulting panel data is highly unbalanced with only 568 individuals observations available for the entire 13 waves. We have decided to keep the dataset intact to control for unobserved characteristics among individuals who tend to complete the surveys, vs those who drop out.

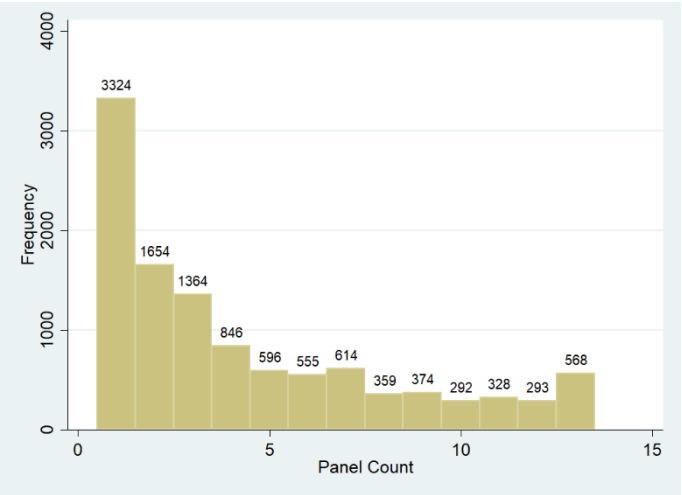

FIGURE 2: DISTRIBUTION OF INDIVIDUALS YEAR WISE RESPONSE TO THE SOCIAL SURVEYS.

4) Econometric Analysis:

The following are the research questions of interest which guide the econometric modelling:
Primary: Does the status of employment of individuals have an impact on their family size(no. of children) ?
Exploratory: Does the gender of the second last child have an impact on family size of an individual ?

4.1) Econometric Analysis on Panel Data:

The core analysis of our research project is exploring the impact of employment status (regressor) on family size (dependent variable). We first run a Poisson regression (Table 5.1) with family size (kids) as the dependent variable and the employment status as regressor. We apply the random effects estimate and cluster by respondent id. The possibility of overdispersion in our regression is eliminated since coefficient of $\hat{y}_{poisson}$**=0.68 < 1.0** (Table 5.2) We then compile the marginal effects of the regression (Table 5.3) to get a better understanding of the impact of the regressor on the dependent variable. Our results indicate that being employed is associated with a higher probability of having more children, but since it is insignificant (**low p value:0.07**), we need to run alternate regressions to confirm the effect.

$$\boldsymbol{Kids_{it} = \beta_0 + \beta_1 \cdot Paid\_Work_{it} + \sum_{j=2008}^{2020} \beta_j \cdot Year_i(dummy) + \varepsilon_{it}}$$

We compare the results of our Poisson regression by running linear regressions on our panel data. We set the family size as dependent variable, employment status as regressor and control for year wise effects, clustering errors by respondent id. The results of our regression (using both random effects (Table 6.1) as well as fixed effects(Table 6.2)) indicate **that there is no strong correlation between family size and employment status**. Our analysis is complicated by the fact that most families tend to 1-3 children and that the impact of the financial crisis on employment status is not captured well by our surveys.

4.2) Expolratory Analysis on Time Invariant Variables:

We have run a linear regression of regressor: education level on dependent family size, with the regressor split into dummy levels, assigning No Schooling as base level (Table 4.0). Our results indicate family size decreases as individuals attain higher education levels (**due to the significant p-values**), which is consistent with our hypothesis.

$$\boldsymbol{Kids_i = \beta_0 \cdot No\_School_i + \sum_{j=2008}^{2020} \beta_j \cdot Educ_i(dummy) + \varepsilon_i}$$

In our exploratory data analysis, we investigate if there is any difference in means of family size by nationality (dutch vs non dutch), gender of second last child (male vs female) and childcare subsidies (yes vs no). The following are the results of the Wilcoxon rank-sum test:

1. There is a significant difference (**p value:0.0016**) between the family size of individuals having their second last child as female vs having their second last child as male. This could indicate to a gender preference among individuals which causes them to try for at least having one male child (Table 4.1)
2. There is no significant difference (**p value:0.195**) between family size of Dutch individuals when compare them with individuals of other nationalities (Table 4.3)
3. Finally we find that individuals with no childcare tend to have larger families (**significant p-value:0.05**) as compared to individuals with no childcare. It is challenging to interpret such a result as we expect childcare allowance to encourage people having more children. It is possible that wealthy individuals are both ineligible for childcare allowance as well as can provide amenities for their children which impacts their decision having larger family size (Table 4.5)

5) Issues and improvements for Further Research

Modelling Issues:

- Since the research question focusses on family size, it would be more appropriate to select the household as a the basic unit of analysis, instead of an individual. It is possible to implement this for future research, by matching individual members with the same household ID. We would also need to match current partners based on household ID, individual ID and other partner characteristics, taking marriage, separation into account (Do you live with your partner? How long have you lived with your partner?, etc).
- It is possible that the employment of the household head has more impact on family size, than the employment status of other household members. It is possible to control for this, in future research, by identifying household heads from the Income Questionnaire (variable: ***001**)
- It is possible that certain female/male respondents voluntarily leave their jobs to raise their family. Thus there could be reverse causality between family size (dependent variable) and employment (regressor).
- There could be unobserved confounding variables impacting family size such as: wealth from non-income sources, health conditions of household members, social beliefs, etc.

Selectivity Issues:

- The survey is conducted in Dutch, hence the sample is skewed towards Dutch speaking individuals living in the Netherlands.
- Many social indicators depend on the validity of the self-reported responses of the participants. There is a high non response rate for many survey questions as well as a requirement to verify the self-reported responses of the participants (regarding income ,education, employment) from government records.
- The unbalanced panel data points to attrition in the sample. There are only 2250 individuals whose data is included in all 13 waves of the survey (2008-2020). This number gets reduced further when we control for age (keeping participants within the age range of 20-60 only).
- The variability of the employment variable (**paid_work**) is low, even in the years of financial crises. This points to a systematic bias of excluding recently unemployed individuals and a need to choose an alternate data source.

6) Tables & Results

Variable	Obs=.	Obs>.	Obs<.	Obs<. Unique values	Obs<. Min	Obs<. Max
id			78,264	>500	800009	899993
year			78,264	13	2008	2020
gender	56		78,208	2	0	1
partner	49		78,215	2	0	1
partner_age	65,164		13,100	93	1810	2006
partner_ge~r	61,683		16,581	2	0	1
kids	9,152		69,112	13	0	15
child_gender	35,129		43,135	2	0	1
childcare	70,123		8,141	2	0	1

Table 1.1: Combined family dataset

Variable	Obs=.	Obs>.	Obs<.	Obs<. Unique values	Obs<. Min	Obs<. Max
id			78,305	>500	800009	899993
year			78,305	13	2008	2020
nationality	15		78,290	2	0	1
abortion	1,486		76,819	4	1	4

Table 1.2: Combined ethnicity dataset

Variable	Obs=.	Obs>.	Obs<.	Obs<. Unique values	Obs<. Min	Obs<. Max
id			72,012	>500	800009	899993
year			72,012	13	2008	2020
fin	9,036		62,976	2	1	2
retire	33,025		38,987	2	0	1

Table 1.3: Combined finance dataset

Variable	Obs=.	Obs>.	Obs<.	Obs<. Unique values	Obs<. Min	Obs<. Max
id			77,227	>500	800009	899993
year			77,227	13	2008	2020
educ	1,296		75,931	5	0	4
paid_work	131		77,096	2	0	1
age			77,227	91	9	109

Table 1.4: Combined work schooling dataset

```
Contains data from combine.dta
  obs:        48,386
 vars:            16                          4 Feb 2022 18:38
 size:     5,806,320
-------------------------------------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------------------------------------
id              double  %10.0g                Number of household member encrypted
year            float   %10.0g                Survey Year
educ            double  %15.0g     educ_level
                                              * What is the highest level of education that you have completed with diploma or c
paid_work       double  %10.0g     cw08a001   Does the respondent have paid work?
age             double  %10.0g                Respondents Age
fin             double  %10.0g     fin_situation
                                              How would you describe the financial situation of your household at this moment?
retire          double  %10.0g     yesno      Were you on early retirement ?
gender          double  %10.0g     sex        Gender respondent
partner         double  %10.0g     yesno      Do you currently have a partner?
partner_age     double  %10.0g                Respondents Partners Age
partner_gender  double  %10.0g     sex        What is your partner's gender?
kids            double  %10.0g                How many children have you had in total?
child_gender    float   %9.0g      sex        Gender of Second Youngest Child
childcare       double  %10.0g     yesno      Have you received any childcare supplement from the tax authority?
nationality     double  %10.0g     yesno      Which language or languages did you grow up speaking: Dutch
abortion        double  %10.0g     yesno      Do you believe that abortion is ever permitted?
                                            * indicated variables have notes
-------------------------------------------------------------------------------------------------------------------------
```

Table 2.1: Final Merged Dataset Description

Variable	Obs	Mean	Std. Dev.	Min	Max
id	48,386	850273	28929.02	800009	899993
year	48,386	2013.572	3.752192	2008	2020
educ	47,796	1.836995	.7431104	0	4
paid_work	48,344	.7948866	.4037886	0	1
age	48,386	42.13473	11.63854	20	60
fin	35,386	1.569802	.4951108	1	2
retire	13,330	.0132783	.1144683	0	1
gender	43,977	.4386156	.4962233	0	1
partner	43,972	.7892295	.4078604	0	1
partner_age	8,347	40.97197	12.31167	15	202
partner_ge~r	10,066	.5659646	.4956542	0	1
kids	37,483	1.597071	1.28278	0	15
child_gender	21,285	.5115339	.4998787	0	1
childcare	7,005	.6301213	.4828061	0	1
nationality	43,250	.9338728	.2485071	0	1
abortion	42,635	1.482303	.8671338	1	4

Table 2.2: Final Merged Dataset Summary Statistics

-> tabulation of nationality

(max) nationality	Freq.	Percent	Cum.
Not Dutch	644	6.35	6.35
Dutch	9,491	93.65	100.00
Total	10,135	100.00	

Table 3.1: Nationality Count Tabulation

-> tabulation of gender

(max) gender	Freq.	Percent	Cum.
female	5,774	55.91	55.91
male	4,553	44.09	100.00
Total	10,327	100.00	

Table 3.2: Gender Count Tabulation

-> tabulation of child_gender

(max) child_gender	Freq.	Percent	Cum.
female	2,340	46.99	46.99
male	2,640	53.01	100.00
Total	4,980	100.00	

Table 3.3: Gender of second last child count Tabulation

-> tabulation of childcare

(max) childcare	Freq.	Percent	Cum.
no	914	35.50	35.50
yes	1,661	64.50	100.00
Total	2,575	100.00	

Table 3.4: Childcare Subsidy Count Tabulation

-> tabulation of educ

(max) educ	Freq.	Percent	Cum.
No Education	69	0.62	0.62
School	3,221	29.06	29.68
University	5,634	50.83	80.51
Advanced Degree	1,608	14.51	95.01
Other	553	4.99	100.00
Total	11,085	100.00	

Table 3.5: Highest education count tabulation

Source	SS	df	MS	Number of obs	=	10,047
				F(4, 10042)	=	46.73
Model	326.367488	4	81.5918719	Prob > F	=	0.0000
Residual	17535.1524	10,042	1.74618128	R-squared	=	0.0183
				Adj R-squared	=	0.0179
Total	17861.5199	10,046	1.77797331	Root MSE	=	1.3214

kids	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ						
School	-.5484902	.1849344	-2.97	0.003	-.9109986	-.1859818
University	-.5402277	.1841665	-2.93	0.003	-.901231	-.1792244
Advanced Degree	-1.014997	.1864691	-5.44	0.000	-1.380513	-.6494799
Other	-.3660377	.1920289	-1.91	0.057	-.7424528	.0103773
_cons	2	.1832496	10.91	0.000	1.640794	2.359206

Table 4.0: Linear Regression of Education Levels on kids

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

child_gender	obs	rank sum	expected
female	2340	5691429	5827770
male	2640	6711261	6574920
combined	4980	12402690	12402690

unadjusted variance 2.564e+09
adjustment for ties -6.957e+08
adjusted variance 1.869e+09

Ho: kids(child_~r==female) = kids(child_~r==male)
z = -3.154
Prob > |z| = 0.0016

Table 4.1: Two Sample Wilcoxon Test by second last child gender

(max) child_gender	Summary of (max) kids Mean	Std. Dev.	Freq.
female	2.5102564	.87888972	2,340
male	2.5640152	.90787916	2,640
Total	2.538755	.89468766	4,980

Table 4.2: Summary Statistics of second last child gender

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

 nationality |       obs    rank sum    expected
-------------+---------------------------------
   Not Dutch |       612     3027352     2944944
       Dutch |      9011    43278524    43360932
-------------+---------------------------------
    combined |      9623    46305876    46305876

unadjusted variance   4.423e+09
adjustment for ties  -3.766e+08
                     ----------
adjusted variance     4.046e+09

Ho: kids(nation~y==Not Dutch) = kids(nation~y==Dutch)
             z =   1.296
    Prob > |z| =   0.1951
```

Table 4.3: Two Sample Wilcoxon Test by Nationality

(max) nationality	Summary of (max) kids Mean	Std. Dev.	Freq.
Not Dutch	1.4869281	1.4308361	612
Dutch	1.3999556	1.3219969	9,011
Total	1.4054869	1.3292744	9,623

Table 4.4: Summary Statistics of Nationality

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

   childcare |       obs    rank sum    expected
-------------+---------------------------------
          no |       914     1326470     1177232
         yes |      1661     1990130     2139368
-------------+---------------------------------
    combined |      2575     3316600     3316600

unadjusted variance   3.259e+08
adjustment for ties   -45638074
                      ----------
adjusted variance     2.803e+08

Ho: kids(childc~e==no) = kids(childc~e==yes)
             z =   8.915
    Prob > |z| =   0.0000
```

Table 4.5: Two Sample Wilcoxon Test by Childcare

(max) childcare	Summary of (max) kids Mean	Std. Dev.	Freq.
no	2.3774617	1.0960327	914
yes	2.0036123	.84024463	1,661
Total	2.1363107	.9557325	2,575

Table 4.6: Summary Statistics of Childcare

```
Fitting Poisson model:

Iteration 0:   log pseudolikelihood = -47083.894
Iteration 1:   log pseudolikelihood = -47083.894

Fitting full model:

Iteration 0:   log pseudolikelihood = -40710.929
Iteration 1:   log pseudolikelihood = -40619.065
Iteration 2:   log pseudolikelihood = -40613.676
Iteration 3:   log pseudolikelihood = -40613.667
Iteration 4:   log pseudolikelihood = -40613.667

Random-effects Poisson regression               Number of obs     =     29,400
Group variable: id                              Number of groups  =      8,058

Random effects u_i ~ Gamma                      Obs per group:
                                                              min =          1
                                                              avg =        3.6
                                                              max =         13

                                                Wald chi2(2)      =    1811.77
Log pseudolikelihood  = -40613.667              Prob > chi2       =     0.0000

                                  (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
       kids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   paid_work |  -.0008189   .0107982    -0.08   0.940    -.021983    .0203451
         fin |  -.0179731   .0066268    -2.71   0.007   -.0309614   -.0049849
       _cons |   .4114377   .0195763    21.02   0.000    .3730689    .4498064
-------------+----------------------------------------------------------------
    /lnalpha |  -.1973198   .1967722                     -.5829862    .1883466
-------------+----------------------------------------------------------------
       alpha |    .820928   .1615358                      .5582289    1.207252
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 1.3e+04                Prob >= chibar2 = 0.000
```

Table 5.1: Panel Data Poisson Regression of Paid Work on Kids

```
. reg ystar yhat_poisson,nocons

      Source |       SS           df       MS      Number of obs   =    37,461
-------------+----------------------------------   F(1, 37460)     =     67.43
       Model |  299.294703         1  299.294703   Prob > F        =    0.0000
    Residual |  166261.127    37,460  4.43836432   R-squared       =    0.0018
-------------+----------------------------------   Adj R-squared   =    0.0018
       Total |  166560.422    37,461  4.44623534   Root MSE        =    2.1067

------------------------------------------------------------------------------
       ystar |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
yhat_poisson |   .0658513   .0080191     8.21   0.000     .0501336    .081569
------------------------------------------------------------------------------
```

Table 5.2: Test for Overdispersion

```
Marginal effects after xtpoisson
      y  = Linear prediction (predict)
         =  .38258931
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.      z    P>|z|  [    95% C.I.   ]       X
---------+--------------------------------------------------------------------
paid_w~k*|  -.0008189       .0108   -0.08   0.940  -.021983  .020345   .803571
     fin |  -.0179731      .00663   -2.71   0.007  -.030961 -.004985   1.56847
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Table 5.3: Marginal Effects of Poisson Regression

```
Fixed-effects (within) regression               Number of obs      =     29,400
Group variable: id                              Number of groups   =      8,058

R-sq:                                           Obs per group:
     within  = 0.0404                                         min =          1
     between = 0.0007                                         avg =        3.6
     overall = 0.0017                                         max =         13

                                                F(14,8057)         =      29.48
corr(u_i, Xb)  = -0.1206                        Prob > F           =     0.0000

                                    (Std. Err. adjusted for 8,058 clusters in id)
------------------------------------------------------------------------------
             |               Robust
        kids |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   paid_work |  -.0169426    .013743    -1.23   0.218    -.0438824    .0099972
         fin |   .0009691   .0088976     0.11   0.913    -.0164725    .0184107
             |
        year |
        2009 |   .1218593   .0092787    13.13   0.000     .1036705     .140048
        2010 |   .1477035    .012573    11.75   0.000     .1230571    .1723498
        2011 |   .2012647   .0123752    16.26   0.000     .1770061    .2255234
        2012 |   .2220341   .0146273    15.18   0.000     .1933608    .2507074
        2013 |   .2612885   .0155709    16.78   0.000     .2307654    .2918115
        2014 |   .2550426   .0167312    15.24   0.000     .2222452    .2878399
        2015 |   .2020026   .0188734    10.70   0.000     .1650059    .2389993
        2016 |   .2371233   .0191942    12.35   0.000     .1994977    .2747489
        2017 |   .2633561   .0204802    12.86   0.000     .2232096    .3035026
        2018 |   .2784498   .0209309    13.30   0.000     .2374198    .3194797
        2019 |   .3075113   .0217449    14.14   0.000     .2648858    .3501369
        2020 |   .3265867   .0225399    14.49   0.000     .2824027    .3707706
             |
       _cons |   1.479267    .020958    70.58   0.000     1.438184     1.52035
-------------+----------------------------------------------------------------
     sigma_u |  1.2780623
     sigma_e |   .3846287
         rho |  .91695257   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Table 6.1 Panel Data Regression (with Fixed Effects)

```
Random-effects GLS regression                   Number of obs     =     29,400
Group variable: id                              Number of groups  =      8,058

R-sq:                                           Obs per group:
     within  = 0.0386                                         min =          1
     between = 0.0005                                         avg =        3.6
     overall = 0.0001                                         max =         13

                                                Wald chi2(14)     =     491.91
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

                                 (Std. Err. adjusted for 8,058 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       kids |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   paid_work |  -.0124699   .0130733    -0.95   0.340    -.0380931    .0131534
         fin |  -.0043302   .0086053    -0.50   0.615    -.0211963    .0125359

        year |
        2009 |   .1523098   .0091996    16.56   0.000     .1342788    .1703407
        2010 |   .1607322   .0123374    13.03   0.000     .1365513     .184913
        2011 |   .2192076   .0120384    18.21   0.000     .1956128    .2428023
        2012 |   .2332987   .0141946    16.44   0.000     .2054777    .2611197
        2013 |    .270672   .0150946    17.93   0.000     .2410871    .3002568
        2014 |   .2511342   .0161372    15.56   0.000     .2195059    .2827624
        2015 |   .1831998   .0180749    10.14   0.000     .1477737    .2186259
        2016 |   .2186924   .0183594    11.91   0.000     .1827086    .2546762
        2017 |   .2390668   .0195426    12.23   0.000      .200764    .2773696
        2018 |   .2505596   .0198013    12.65   0.000     .2117499    .2893694
        2019 |   .2810756   .0206086    13.64   0.000     .2406835    .3214676
        2020 |   .2897881    .021173    13.69   0.000     .2482898    .3312864

       _cons |   1.259359   .0239929    52.49   0.000     1.212334    1.306385
-------------+----------------------------------------------------------------
     sigma_u |  1.1433578
     sigma_e |   .3846287
         rho |   .89833797   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Table 6.2: Panel Data Regression (with Random Effects)