Консультация к КР.

1. Пусть $y_t^* = x_t'\beta + \varepsilon_t$, где ошибки ε_t имеют плотность распределения $f(x)$ и

$$y_t = \begin{cases} \alpha_1, & \text{если } y_t^* \leqslant \alpha_1 \\ y_t^*, & \text{если } \alpha_1 < y_t^* < \alpha_2 \\ \alpha_2, & \text{если } y_t^* \geqslant \alpha_2 \end{cases}$$

   а) Найдите распределение y_t.

   б) Найдите логарифмическую функцию правдоподобия для оценивания вектора β.

   в) Найдите $(\partial \mathrm{E} y)/(\partial x)$.

2. Рассмотрим модель для панельных данных:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + u_{it},\ i = 1, ..., N,\ t = 1, ..., T.$$

   Обозначим $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\varepsilon_{it} = \alpha_i + u_{it}$. Рассмотрим преобразованную модель:

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1\left(x_{it} - \lambda\bar{x}_i\right) + \left(\varepsilon_{it} - \lambda\bar{\varepsilon}_i\right).$$

   (a) Какие модели получаются при $\lambda = 0$ и при $\lambda = 1$?

   (b) Пусть $\alpha_i \sim$ i.i.d. $(0, \sigma_\alpha^2)$; $u_{it} \sim$ i.i.d. $(0, \sigma_u^2)$; $\mathrm{Cov}\left(\alpha_i, u_{it}\right) = 0$ для всех i и j. Определим

$$\lambda = 1 - \left[\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}\right]^{1/2}.$$

   Покажите, что $\varepsilon_{it} - \lambda\bar{\varepsilon}_i$ имеет нулевое матетматическое ожидание, постоянную дисперсию и серийная корреляция отсутствует.

3. Рассмотрим модель с фиксированными эффектами:

$$y_{it} = \alpha_i + \varepsilon_{it},\ i = 1, ..., n,\ t = 1, ..., T.$$

   Случайные ошибки предполагаются независимыми и гетероскедастичными, то есть $V\left(\varepsilon_{it}\right) = \sigma_i^2$. Панель является несбалансированной, то есть каждому i-му субъекту в выборке соответствуют T_i наблюдений.

   (a) Покажите, что OLS и GLS оценки α_i совпадают.

   (b) Пусть $\sigma^2 = \sum_{i=1}^{N} T_i \frac{\sigma_i^2}{n}, n = \sum_{i=1}^{N} T_i$ — дисперсия взвешенной случайной ошибки. Покажите, что OLS оценка для σ^2 является смещенной. Также покажите, что смещение исчезает, если панель сбалансированная и случайные ошибки гомоскедастичны.

Problem 1.

The data set contains data on arrests during 1986 on 2,725 men born in California in 1960 or 1961. Each man in the sample was arrested at least once prior to 1986. We are interested in explaining what determines how often these men are arrested again in 1986. The following variables are available.

narr86	# times arrested, 1986
avgsen	average sentence length served for prior convictions (in months)
black	=1 if black
born60	=1 if born in 1960
durat	recent unemployment duration (in months)
hispan	=1 if Hispanic
inc86	legal income, 1986, $100s
pcnv	proportion of prior convictions
ptime86	months in prison during 1986
qemp86	# quarters employed, 1986
tottime	time in prison since 18 (in months)

The variable *pcnv*, the proportion of arrests prior to 1986 that led to conviction, is a proxy for the likelihood of being convicted for a crime.
The variable *avgsen* is a proxy for the severity of punishment, if convicted.
The variable *ptime86* captures the incarcerative effect of crime: if an individual is in prison, he cannot be arrested for a crime outside of prison. Labour market opportunities are captured by *qemp86*.

Summary statistics:

	avgsen	black	born60	durat	hispan	inc86	pcnv	ptime86	qemp86	tottime
Mean	0.632	0.161	0.363	2.251	0.218	54.97	0.358	0.387	2.309	0.839
Median	0	0	0	0	0	29	0.25	0	3	0
Maximum	59.2	1	1	25	1	541	1	12	4	63.4
Minimum	0	0	0	0	0	0	0	0	0	0
Std. Dev.	3.508	0.368	0.481	4.607	0.413	66.63	0.395	1.950	1.610	4.607

Tabulation of narr86

Value	Count	Percent	Cumulative Count	Cumulative Percent
0	1970	72.29	1970	72.29
1	559	20.51	2529	92.81
2	121	4.44	2650	97.25
3	42	1.54	2692	98.79
4	12	0.44	2704	99.23
5	13	0.48	2717	99.71
6	4	0.15	2721	99.85
7	1	0.04	2722	99.89
9	1	0.04	2723	99.93
10	1	0.04	2724	99.96
12	1	0.04	2725	100.00
Total	2725	100.00	2725	100.00

Three models were estimated:

Dependent Variable: NARR86>0 **MODEL 1**
Method: ML - Binary Probit (Quadratic hill climbing)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	−0.323473	0.057875	−5.589133	0.0000
PCNV	−0.553008	0.071956	−7.685374	0.0000
AVGSEN	0.003730	0.007684	0.485502	0.6273

PTIME86	–0.081029	0.018028	–4.494518	0.0000
QEMP86	0.010286	0.024660	0.417109	0.6766
INC86	–0.004832	0.000686	–7.045535	0.0000
BLACK	0.467772	0.072006	6.496325	0.0000
HISPAN	0.288913	0.065449	4.414352	0.0000

Mean dependent var	0.277064	S.D. dependent var	0.447631
S.E. of regression	0.428541	Akaike info criterion	1.094808
Sum squared resid	498.9705	Schwarz criterion	1.112160
Log likelihood	–1483.677	Hannan–Quinn criter.	1.101080
Restr. log likelihood	–1608.184	Avg. log likelihood	–0.544468
LR statistic (7 df)	249.0143	McFadden R–squared	0.077421
Probability(LR stat)	0.000000		
Obs with Dep=0	1970	Total obs	2725
Obs with Dep=1	755		

MODEL 2

Method: ML - Ordered Logit (Quadratic hill climbing)
Included observations: 2725
Number of ordered indicator values: 11

	Coefficient	Std. Error	z-Statistic	Prob.
PCNV	–0.826305	0.121661	–6.791867	0.0000
AVGSEN	0.009250	0.012395	0.746213	0.4555
PTIME86	–0.132007	0.030301	–4.356456	0.0000
QEMP86	0.004806	0.041021	0.117167	0.9067
INC86	–0.008754	0.001267	–6.907944	0.0000
BLACK	0.797965	0.114772	6.952618	0.0000
HISPAN	0.533740	0.108032	4.940569	0.0000
	Limit Points			
LIMIT_1:C(8)	0.518862	0.094332	5.500371	0.0000
LIMIT_2:C(9)	2.213780	0.111413	19.87008	0.0000
LIMIT_3:C(10)	3.242495	0.143630	22.57529	0.0000
LIMIT_4:C(11)	4.085995	0.193895	21.07326	0.0000
LIMIT_5:C(12)	4.544892	0.234307	19.39718	0.0000
LIMIT_6:C(13)	5.518044	0.363691	15.17233	0.0000
LIMIT_7:C(14)	6.213840	0.507219	12.25080	0.0000
LIMIT_9:C(15)	6.502190	0.583614	11.14125	0.0000
LIMIT_10:C(16)	6.908442	0.712230	9.699735	0.0000
LIMIT_12:C(17)	7.602260	1.003628	7.574779	0.0000

Akaike info criterion	1.582839	Schwarz criterion	1.619710
Log likelihood	–2139.618	Hannan–Quinn criter.	1.596166
Restr. log likelihood	–2269.057	Avg. log likelihood	–0.785181
LR statistic (7 df)	258.8775	LR index (Pseudo–R2)	0.057045
Probability(LR stat)	0.000000		

Dependent Variable: NARR86 **MODEL 3**
Method: ML/QML - Poisson Count (Quadratic hill climbing)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	–0.616217	0.063618	–9.686227	0.0000
PCNV	–0.401797	0.084783	–4.739146	0.0000
AVGSEN	0.005603	0.007437	0.753388	0.4512
PTIME86	–0.093507	0.020314	–4.603135	0.0000
QEMP86	–0.037305	0.028907	–1.290529	0.1969
INC86	–0.008127	0.001037	–7.833997	0.0000
BLACK	0.661036	0.073871	8.948584	0.0000
HISPAN	0.501855	0.073884	6.792448	0.0000

R–squared	0.075108	Mean dependent var	0.404404

Adjusted R–squared	0.072725	S.D. dependent var	0.859077
S.E. of regression	0.827249	Akaike info criterion	1.657495
Sum squared resid	1859.353	Schwarz criterion	1.674846
Log likelihood	–2250.337	Hannan–Quinn criter.	1.663767
Restr. log likelihood	–2441.921	Avg. log likelihood	–0.825812
LR statistic (7 df)	383.1683	LR index (Pseudo–R2)	0.078456
Probability(LR stat)	0.000000		

(a) (6 points) Use each of three models to estimate probability to be unarrested, $P0$, in 1986 for the two men, with values of the variables: avgsen = 0, inc86 = 50, pcnv = 0.1, ptime86 = 0, qemp86 = 3, one of which is white and another is black.

(b) (6 points) Use each of three models to estimate marginal effect of income, $\frac{\partial P0}{\partial\, inc86}$ for the black man in 1).

(c) (4 points) Comment the models results. Which variables affect the crime (number of arrests)? Discuss some findings that you think are interesting.

(d) (4 points) Discuss shortly limitations and advantages of these models. Which model(s) is more appropriate to that problem in your opinion?

Problem 2.

По панели 78 российских регионов за 9 лет оценивается три спецификации модели для миграции в зависимости от лагов следующих переменных: человеческого капитала (L1x1), уровня безработицы (L1x2), логарифма среднедушевого дохода (L1x3), численности студентов на душу населения (L1student_pc), коэффициента Джини (L1gini), доли молодого населения (L1young), доли пожилого населения (L1old). В модели также учен временной эффект в виде набора дамми переменных на года (year7-year10).

Прокомментируйте, какая модель:

- сквозная (pooled) регрессия;
- модель с фиксированными эффектами;
- модель со случайными эффектами

должна быть выбрана в данном случае. В своих рассуждениях приводите в подтверждение результаты соответствующих тестов, в каждом из тестов опишите основную и альтернативную гипотезы. Ниже приведены результаты оценивания моделей сквозной регрессии, с индивидуальными эффектами — фиксированными и случайными — а также результаты тестов на выбор спецификации модели.

Модель 1.

```
. reg net_mig_all L1* year2007-year2010

      Source |       SS       df       MS              Number of obs =     702
-------------+------------------------------           F( 12,   689) =   44.41
       Model |   20210.316    12    1684.193           Prob > F      =  0.0000
    Residual |  26131.1163   689  37.9261485           R-squared     =  0.4361
-------------+------------------------------           Adj R-squared =  0.4263
       Total |  46341.4324   701  66.1076068           Root MSE      =  6.1584

------------------------------------------------------------------------------
 net_mig_all |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        L1x1 |   -.075536   .0555915    -1.36   0.175     -.184685    .0336131
        L1x2 |  -1.679025   .7732123    -2.17   0.030     -3.19716     -.16089
        L1x3 |   6.087226   1.488182     4.09   0.000     3.165311    9.009142
        L1x4 |   2.756517   1.136344     2.43   0.016     .5254044     4.98763
L1student_pc |   .1215523   .0195059     6.23   0.000     .0832542    .1598504
      L1gini |   -12.1918   11.55797    -1.05   0.292    -34.88487    10.50128
     L1young |   1.266557   .1590198     7.96   0.000      .954335    1.578778
       L1old |   1.722642   .1046239    16.47   0.000     1.517222    1.928061
    year2007 |  -.0010123   .8082426    -0.00   0.999    -1.587926    1.585902
    year2008 |  -1.468292   .8146773    -1.80   0.072     -3.06784    .1312557
    year2009 |  -1.661877   .8457649    -1.96   0.050    -3.322463   -.0012914
    year2010 |  -3.477737   .9235795    -3.77   0.000    -5.291105   -1.664369
       _cons |   -60.8354   6.263506    -9.71   0.000    -73.13325   -48.53755
------------------------------------------------------------------------------
```

Модель 2.

```
. xtreg net_mig_all L1* year2007-year2010, fe

Fixed-effects (within) regression               Number of obs      =       702
Group variable: region                          Number of groups   =        78

R-sq:  within  = 0.2129                         Obs per group: min =         9
       between = 0.4225                                        avg =       9.0
       overall = 0.3924                                        max =         9

                                                F(12,612)          =     13.79
corr(u_i, Xb)  = 0.2382                         Prob > F           =    0.0000

------------------------------------------------------------------------------
 net_mig_all |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        L1x1 |  -.0600761   .0461687    -1.30   0.194    -.1507444    .0305921
        L1x2 |  -1.695126   .6243793    -2.71   0.007    -2.921312   -.4689405
        L1x3 |   7.313121   1.470259     4.97   0.000     4.425756    10.20049
        L1x4 |   4.384781   .5188239     8.45   0.000      3.36589    5.403672
L1student_pc |   .0735784   .0311219     2.36   0.018     .0124598    .1346971
      L1gini |  -30.41007   12.19753    -2.49   0.013    -54.36415   -6.455984
     L1young |   1.366945   .2380267     5.74   0.000     .8994971    1.834393
       L1old |   1.335605   .2752427     4.85   0.000     .7950704     1.87614
    year2007 |   .4974565    .423644     1.17   0.241    -.3345157    1.329429
    year2008 |  -.6770634   .5033005    -1.35   0.179    -1.665469    .3113422
    year2009 |  -.6682617   .5620948    -1.19   0.235     -1.77213    .4356069
    year2010 |  -2.148101   .6736989    -3.19   0.002    -3.471143   -.8250588
       _cons |  -49.47221    8.73733    -5.66   0.000      -66.631   -32.31343
-------------+----------------------------------------------------------------
     sigma_u |  6.0222964
     sigma_e |  2.6959466
         rho |  .83305546   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(77, 612) =    38.74             Prob > F = 0.0000

. est store fe
```

Модель 3.

```
. xtreg net_mig_all L1* year2007-year2010, re

Random-effects GLS regression                   Number of obs      =       702
Group variable: region                          Number of groups   =        78
R-sq:  within  = 0.2104                         Obs per group: min =         9
       between = 0.4562                                        avg =       9.0
       overall = 0.4258                                        max =         9
Random effects u_i ~ Gaussian                   Wald chi2(12)      =    225.25
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
 net_mig_all |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        L1x1 |  -.0512574     .04406    -1.16   0.245    -.1376134    .0350985
        L1x2 |  -1.698741    .602328    -2.82   0.005    -2.879282   -.5181995
        L1x3 |   6.720255   1.358439     4.95   0.000     4.057763    9.382746
        L1x4 |   4.308914   .5173729     8.33   0.000     3.294882    5.322946
L1student_pc |    .085199   .0255655     3.33   0.001     .0350915    .1353065
      L1gini |  -24.92428   11.29259    -2.21   0.027    -47.05735   -2.791216
     L1young |    1.28032    .194049     6.60   0.000     .8999912    1.660649
       L1old |   1.641813   .1616724    10.16   0.000     1.324941    1.958685
    year2007 |   .2547917   .3985384     0.64   0.523    -.5263292    1.035913
    year2008 |  -1.077568   .4439989    -2.43   0.015     -1.94779   -.2073465
    year2009 |  -1.145741   .4851141    -2.36   0.018    -2.096548   -.1949351
    year2010 |  -2.738704   .5645497    -4.85   0.000    -3.845201   -1.632207
       _cons |  -56.10931   7.501452    -7.48   0.000    -70.81188   -41.40673
-------------+----------------------------------------------------------------
     sigma_u |  5.6929442
     sigma_e |  2.6959466
         rho |  .81682103   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. est store re
. xttest0
Breusch and Pagan Lagrangian multiplier test for random effects
        net_mig_all[region,t] = Xb + u[region] + e[region,t]
        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
               net_mig~l |   66.10761       8.130658
                       e |   7.268128       2.695947
                       u |   32.40961       5.692944

        Test:   Var(u) = 0
                             chi2(1) =  1808.88
                          Prob > chi2 =     0.0000

. hausman fe re
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
-------------+----------------------------------------------------------------
        L1x1 |   -.0600761    -.0512574       -.0088187        .0137937
        L1x2 |   -1.695126    -1.698741        .0036144        .1644702
        L1x3 |    7.313121     6.720255        .5928663        .5624094
        L1x4 |    4.384781     4.308914        .0758666        .0387748
L1student_pc |    .0735784      .085199       -.0116206        .0177475
      L1gini |   -30.41007    -24.92428       -5.485783        4.610537
     L1young |    1.366945      1.28032        .0866249        .1378465
       L1old |    1.335605     1.641813       -.3062076        .2227568
    year2007 |    .4974565     .2547917        .2426649        .1436711
    year2008 |   -.6770634    -1.077568        .4005049        .2370156
    year2009 |   -.6682617    -1.145741        .4774796        .2839276
    year2010 |   -2.148101    -2.738704        .5906032        .3676327
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic
```

```
 chi2(12) = (b-B)'[(V_b-V_B)^(-1)](b-B)
          =        6.54
Prob>chi2 =      0.8863
```