Эконометрика, 2017-2018, 1 модуль

Семинары 1 - 2

2.04.18 и 9.04.18 для

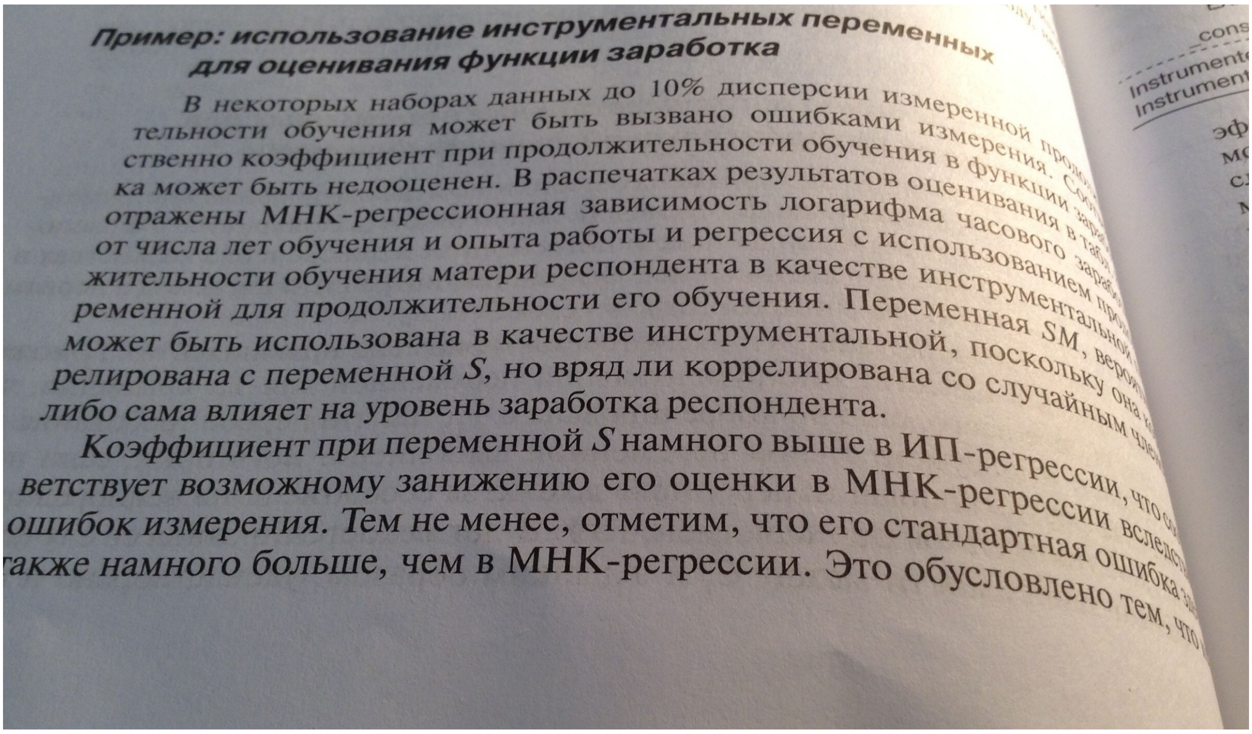
Группы Э_Б2015_Э_3

Семинарист О.А.Демидова

Критика М.Фридменом стандартной функции потребления, раздел 8.5.

1) (Доугерти, 8.7) В некоторой экономике дисперсия переменного дохода составляет 0.5 от дисперсии постоянного дохода, склонность к потреблению товаров кратковременного пользования за счет постоянного дохода составляет 0.6, а расходы на товары длительного пользования отсутствуют. Каким будет значение мультипликатора, полученного на оснве построения «наивной» регрессионной зависимости потребления от дохода, и каково его истинное значение?

2) (Доугерти, раздел 8)



Пример: использование инструментальных переменных для оценивания функции заработка

В некоторых наборах данных до 10% дисперсии измеренной продол-
тельности обучения может быть вызвано ошибками измерения. Соот-
ственно коэффициент при продолжительности обучения в функции зар-
ка может быть недооценен. В распечатках результатов оценивания в таб-
отражены МНК-регрессионная зависимость логарифма часового зараб-
от числа лет обучения и опыта работы и регрессия с использованием прод-
жительности обучения матери респондента в качестве инструментальной
ременной для продолжительности его обучения. Переменная SM, веро-
может быть использована в качестве инструментальной, поскольку она к-
релирована с переменной S, но вряд ли коррелирована со случайным чле-
либо сама влияет на уровень заработка респондента.

Коэффициент при переменной S намного выше в ИП-регрессии, что со-
ветствует возможному занижению его оценки в МНК-регрессии вследст-
ошибок измерения. Тем не менее, отметим, что его стандартная ошибка за-
также намного больше, чем в МНК-регрессии. Это обусловлено тем, что

Таблица 8.2

```
reg LGEARN S EXP

  Source |       SS           df       MS              Number of obs =     540
---------+------------------------------               F(2,537)      =  100.86
   Model |  50.9842581         2   25.492129           Prob > F      =  0.0000
Residual |  135.723385       537  .252743734           R-squared     =  0.2731
---------+------------------------------               Adj R-squared =  0.2704
   Total |  186.707643       539  .34639637            Root MSE      =  .50274

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .1235911   .0090989    13.58   0.000     .1057173     .141465
     EXP |   .0350826   .0050046     7.01   0.000     .0252515    .0449137
   _cons |   .5093196   .1663823     3.06   0.002     .1824796    .8361596
------------------------------------------------------------------------------

ivreg LGEARN EXP (S=SM)
Instrumental variables (2SLS) regression

  Source |       SS           df       MS              Number of obs =     540
---------+------------------------------               F(2,537)      =   28.38
   Model |  46.9446075         2  23.4723038           Prob > F      =  0.0000
Residual |  139.763036       537  .260266361           R-squared     =  0.2514
---------+------------------------------               Adj R-squared =  0.2486
   Total |  186.707643       539  .34639637            Root MSE      =  .51016

------------------------------------------------------------------------------
  LGEARN |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
       S |   .1599676   .0252801     6.33   0.000     .1103076    .2096277
     EXP |   .0394422   .0058092     6.79   0.000     .0280306    .0508537
   _cons |  -.0617062   .4061769    -0.15   0.879    -.8595966    .7361841
------------------------------------------------------------------------------
Instrumented: S
Instruments:  EXP SM
```

…ключ «со…
…бодный член име…
…в данном случае смысл од…
…поэтому константа включена.

Таблица 8.4

. ivreg LGEARN EXP ASVABC MALE ETHBLACK ETHHISP (S=SM SF SIBLINGS LIBRARY)

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	64.4915831	6	10.7485972
Residual	122.21606	533	.229298424
Total	186.707643	539	.34639637

Number of obs	=	540
F(6,533)	=	37.…
Prob > F	=	0.0…
R-squared	=	0.3…
Adj R-squared	=	0.3…
Root MSE	=	.47…

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.111379	.0476886	2.34	0.020	.0176984	[illegible]
EXP	.0258798	.0081187	3.19	0.002	.0099313	[illegible]
ASVABC	.0092263	.007991	1.15	0.249	−.0064714	[illegible]
MALE	.2619787	.0429283	6.10	0.000	.1776492	[illegible]
ETHBLACK	−.0121846	.0822942	−0.15	0.882	−.1738454	[illegible]
ETHHISP	.0457639	.0955115	0.48	0.632	−.1418612	[illegible]
_cons	.2258512	.3887468	0.58	0.562	−.5378125	[illegible]

Instrumented: S

Instruments: EXP ASVABC MALE ETHBLACK ETHHISP SM SF SIBLINGS LIBRARY

. estimates store EARNIV

. reg LGEARN S EXP ASVABC MALE ETHBLACK ETHHISP

Source	SS	df	MS
Model	65.490707	6	10.9151178
Residual	121.216936	533	.227423895
Total	186.707643	539	.34639637

Number of obs	=	540
F(6,533)	=	47.9…
Prob > F	=	0.000…
R-squared	=	0.35…
Adj R-squared	=	0.34…
Root MSE	=	.47…

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.0883257	.0109987	8.03	0.000	.0667196	[illegible]
EXP	.0227131	.0050095	4.53	0.000	.0128724	[illegible]
ASVABC	0129274	.0028834	4.48	0.000	.0072633	[illegible]
MALE	.2652878	.042235	6.28	0.000	.1823203	[illegible]
ETHBLACK	.0077265	.0715863	0.11	0.914	−.1328994	[illegible]
ETHHISP	.0536544	.0937966	0.57	0.568	−.1306019	[illegible]
_cons	.4002952	.1663149	2.41	0.016	.0735821	[illegible]

estimates store EARNOLS
hausman EARNIV EARNOLS, constant

— Coefficients —

	(b) EARNIV	(B) EARNOLS	(b – B) Difference	sqrt(diag (V_b-V_B)) S.E.
S	.111379	.0883257	.0230533	.0464029
EXP	.0258798	.0227131	.0031667	.0063889
ASVABC	.0092263	.0129274	–.0037011	.0074527
MALE	.2619787	.2652878	–.0033091	.0076842
ETHBLACK	–.0121846	.0077265	–.019911	.0405924
ETHHISP	.0457639	.0536544	–.0078904	.018018
_cons	.2258512	.4002952	–.174444	.3513736

b = consistent under Ho and Ha; obtained from ivreg
B = inconsistent under Ha, efficient under Ho; obtained from regress

Test: Ho: difference in coefficients not systematic

chi2(7) = (b – B)'[(V_b – V_B)^(–1)](b – B) = = 0.25

Prob>chi2 = 0.9999

3) Cameron, Trivedy, Microeconometrics using STATA

6.3.2 Medical expenditures with one endogenous regressor

We consider a model with one endogenous regressor, several exogenous regressors, and one or more excluded exogenous variables that serve as the identifying instruments.

The dataset is an extract from the Medical Expenditure Panel Survey (MEPS) of individuals over the age of 65 years, similar to the dataset described in section 3.2.1. The equation to be estimated has the dependent variable `ldrugexp`, the log of total out-of-pocket expenditures on prescribed medications. The regressors are an indicator for whether the individual holds either employer or union-sponsored health insurance (`hi_empunion`), number of chronic conditions (`totchr`), and four sociodemographic variables: age in years (`age`), indicators for whether female (`female`) and whether black or Hispanic (`blhisp`), and the natural logarithm of annual household income in thousands of dollars (`linc`).

We treat the health insurance variable `hi_empunion` as endogenous. The intuitive justification is that having such supplementary insurance on top of the near universal Medicare insurance for the elderly may be a choice variable. Even though most individuals in the sample are no longer working, those who expected high future medical expenses might have been more likely to choose a job when they were working that would provide supplementary health insurance upon retirement. Note that Medicare did not cover drug expenses for the time period we study.

We use the global macro `x2list` to store the names of the variables that are treated as exogenous regressors. We have

```
. * Read data, define global x2list, and summarize data
. use mus06data.dta

. global x2list totchr age female blhisp linc

. summarize ldrugexp hi_empunion $x2list
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ldrugexp	10391	6.479668	1.363395	0	10.18017
hi_empunion	10391	.3796555	.4853245	0	1
totchr	10391	1.860745	1.290131	0	9
age	10391	75.04639	6.69368	65	91
female	10391	.5797325	.4936256	0	1
blhisp	10391	.1703397	.3759491	0	1
linc	10089	2.743275	.9131433	-6.907755	5.744476

6.3.3 Available instruments

We consider four potential instruments for hi_empunion. Two reflect the income status of the individual and two are based on employer characteristics.

The ssiratio instrument is the ratio of an individual's social security income to the individual's income from all sources, with high values indicating a significant income constraint. The lowincome instrument is a qualitative indicator of low-income status. Both these instruments are likely to be relevant, because they are expected to be negatively correlated with having supplementary insurance. To be valid instruments, we need to assume they can be omitted from the equation for ldrugexp, arguing that the direct role of income is adequately captured by the regressor linc.

The firmsz instrument measures the size of the firm's employed labor force, and the multlc instrument indicates whether the firm is a large operator with multiple locations. These variables are intended to capture whether the individual has access to supplementary insurance through the employer. These two variables are irrelevant for those who are retired, self-employed, or purchase insurance privately. In that sense, these two instruments could potentially be weak.

```
. * Summarize available instruments
. summarize ssiratio lowincome multlc firmsz if linc!=.

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    ssiratio |     10089    .5365438    .3678175          0    9.25062
   lowincome |     10089    .1874319    .3902771          0          1
      multlc |     10089    .0620478    .2412543          0          1
      firmsz |     10089    .1405293    2.170389          0         50
```

We have four available instruments for one endogenous regressor. The obvious approach is to use all available instruments, because in theory this leads to the most efficient estimator. In practice, it may lead to larger small-sample bias because the small-sample biases of IV estimators increase with the number of instruments (Hahn and Hausman 2002).

At a minimum, it is informative to use correlate to view the gross correlation between endogenous variables and instruments and between instruments. When multiple instruments are available, as in the case of overidentified models, then it is actually the partial correlation after controlling for other available instruments that matters. This important step is deferred to sections 6.4.2 and 6.4.3.

6.3.4 IV estimation of an exactly identified model

We begin with IV regression of `ldrugexp` on the endogenous regressor `hi_empunion`, instrumented by the single instrument `ssiratio`, and several exogenous regressors.

We use `ivregress` with the `2sls` estimator and the options `vce(robust)` to control for heteroskedastic errors and `first` to provide output that additionally reports results from the first-stage regression. The output is in two parts:

```
. * IV estimation of a just-identified model with single endog regressor
. ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, vce(robust) first

First-stage regressions
-----------------------

                                                Number of obs   =      10089
                                                F(   6,  10082) =     119.18
                                                Prob > F        =     0.0000
                                                R-squared       =     0.0761
                                                Adj R-squared   =     0.0755
                                                Root MSE        =     0.4672

------------------------------------------------------------------------------
             |               Robust
 hi_empunion |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      totchr |   .0127865   .0036655     3.49   0.000     .0056015    .0199716
         age |  -.0086323   .0007087   -12.18   0.000    -.0100216   -.0072431
      female |   -.07345    .0096392    -7.62   0.000    -.0923448   -.0545552
      blhisp |   -.06268    .0122742    -5.11   0.000      -.08674   -.0386201
        linc |   .0483937   .0066075     7.32   0.000     .0354417    .0613456
    ssiratio |  -.1916432   .0236326    -8.11   0.000    -.2379678   -.1453186
       _cons |   1.028981   .0581387    17.70   0.000     .9150172    1.142944
------------------------------------------------------------------------------

Instrumental variables (2SLS) regression               Number of obs =   10089
                                                       Wald chi2(6)  = 2000.86
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.0640
                                                       Root MSE      =  1.3177

------------------------------------------------------------------------------
             |               Robust
    ldrugexp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 hi_empunion |  -.8975913   .2211268    -4.06   0.000    -1.330992   -.4641908
      totchr |   .4502655   .0101969    44.16   0.000       .43028     .470251
         age |  -.0132176   .0029977    -4.41   0.000    -.0190931   -.0073421
      female |   -.020406   .0326114    -0.63   0.531    -.0843232    .0435113
      blhisp |  -.2174244   .0394944    -5.51   0.000     -.294832   -.1400167
        linc |   .0870018   .0226356     3.84   0.000     .0426368    .1313668
       _cons |    6.78717   .2688453    25.25   0.000     6.260243    7.314097
------------------------------------------------------------------------------
Instrumented:  hi_empunion
Instruments:   totchr age female blhisp linc ssiratio
```

6.3.6 Testing for regressor endogeneity

The preceding analysis treats the insurance variable, `hi_empunion`, as endogenous. If instead the variable is exogenous, then the IV estimators (IV, 2SLS, or GMM) are still consistent, but they can be much less efficient than the OLS estimator.

The Hausman test principle provides a way to test whether a regressor is endogenous. If there is little difference between OLS and IV estimators, then there is no need to instrument, and we conclude that the regressor was exogenous. If instead there is considerable difference, then we needed to instrument and the regressor is endogenous. The test usually compares just the coefficients of the endogenous variables. In the case of just one potentially endogenous regressor with a coefficient denoted by β, the Hausman test statistic

$$T_H = \frac{(\widehat{\beta}_{\text{IV}} - \widehat{\beta}_{\text{OLS}})^2}{\widehat{V}(\widehat{\beta}_{\text{IV}} - \widehat{\beta}_{\text{OLS}})}$$

is $\chi^2(1)$ distributed under the null hypothesis that the regressor is exogenous.

Before considering implementation of the test, we first obtain the OLS estimates to compare them with the earlier IV estimates. We have

```
. * Obtain OLS estimates to compare with preceding IV estimates
. regress ldrugexp hi_empunion $x2list, vce(robust)

Linear regression                                      Number of obs =   10089
                                                       F(  6, 10082)    876.85
                                                       Prob > F         0.0000
                                                       R-squared        0.1770
                                                       Root MSE         =1.236
```

ldrugexp	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
hi_empunion	.0738788	.0259848	2.84	0.004	.0229435	.1248141
totchr	.4403807	.0093633	47.03	0.000	.4220268	.4587346
age	-.0035295	.001937	-1.82	0.068	-.0073264	.0002675
female	.0578055	.0253651	2.28	0.023	.0080848	.1075262
blhisp	-.1513068	.0341264	-4.43	0.000	-.2182013	-.0844122
linc	.0104815	.0137126	0.76	0.445	-.0163979	.037361
_cons	5.861131	.1571037	37.31	0.000	5.553176	6.169085

The OLS estimates differ substantially from the just-identified IV estimates given in section 6.3.4. The coefficient of hi_empunion has an OLS estimate of 0.074, greatly different from the IV estimate of −0.898. This is strong evidence that hi_empunion is endogenous. Some coefficients of exogenous variables also change, notably, those for age and female. Note also the loss in precision in using IV. Most notably, the standard error of the instrumented regressor increases from 0.026 for OLS to 0.221 for IV, an eightfold increase, indicating the potential loss in efficiency due to IV estimation.

The hausman command can be used to compute T_H under the assumption that $\widehat{V}(\widehat{\beta}_{\rm IV}-\widehat{\beta}_{\rm OLS}) = \widehat{V}(\widehat{\beta}_{\rm IV}) - \widehat{V}(\widehat{\beta}_{\rm OLS})$; see section 12.7.5. This greatly simplifies analysis because then all that is needed are coefficient estimates and standard errors from separate IV estimation (IV, 2SLS, or GMM) and OLS estimation. But this assumption is too strong. It is correct only if $\widehat{\beta}_{\rm OLS}$ is the fully efficient estimator under the null hypothesis of exogeneity, an assumption that is valid only under the very strong assumption that model errors are independent and homoskedastic. One possible variation is to perform an appropriate bootstrap; see section 13.4.6.

The postestimation estat endogenous command implements the related Durbin–Wu–Hausman (DWH) test. Because the DWH test uses the device of augmented regressors, it produces a robust test statistic (Davidson 2000). The essential idea is the following. Consider the model as specified in section 6.2.1. Rewrite the structural equation (6.2) with an additional variable, v_1, that is the error from the first-stage equation (6.3) for y_2. Then

$$y_{1i} = \beta_1 y_{2i} + \mathbf{x}'_{1i}\beta_2 + \rho v_{1i} + u_i$$

Under the null hypothesis that y_2 is exogenous, $E(v_{1i}u_i|y_{2i}, \mathbf{x}_{1i}) = 0$. If v_1 could be observed, then the test of exogeneity would be the test of $H_0: \rho = 0$ in the OLS regression of y_1 on y_2, $\mathbf{x}_1$, and v_1. Because v_1 is not directly observed, the fitted residual vector

$\hat{v}_1$ from the first-stage OLS regression (6.3) is instead substituted. For independent homoskedastic errors, this test is asymptotically equivalent to the earlier Hausman test. In the more realistic case of heteroskedastic errors, the test of $H_0 : \rho = 0$ can still be implemented provided that we use robust variance estimates. This test can be extended to the multiple endogenous regressors case by including multiple residual vectors and testing separately for correlation of each with the error on the structural equation.

We apply the test to our example with one potentially endogenous regressor, hi_empunion, instrumented by ssiratio. Then

```
. * Robust Durbin-Wu-Hausman test of endogeneity implemented by estat endogenous
. ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, vce(robust)

Instrumental variables (2SLS) regression               Number of obs =    10089
                                                       Wald chi2(6)     2000.86
                                                       Prob > chi2       0.0000
                                                       R-squared         0.0640
                                                       Root MSE          1.8177

------------------------------------------------------------------------------
             |               Robust
    ldrugexp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 hi_empunion |  -.8975913   .2211268    -4.06   0.000    -1.330992   -.4641908
      totchr |   .4502655   .0101969    44.16   0.000       .43028     .470251
         age |  -.0132176   .0029977    -4.41   0.000    -.0190931   -.0073421
      female |   -.020406   .0326114    -0.63   0.531    -.0843232    .0435113
      blhisp |  -.2174244   .0394944    -5.51   0.000     -.294832   -.1400167
       linc |   .0870018   .0226356     3.84   0.000     .0426368    .1313668
       _cons |    6.78717   .2688453    25.25   0.000     6.260243    7.314097
------------------------------------------------------------------------------
Instrumented:  hi_empunion
Instruments:   totchr age female blhisp linc ssiratio

. estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Robust score chi2(1)            =  24.935  (p = 0.0000)
  Robust regression F(1,10081)    = 26.4333  (p = 0.0000)
```

The last line of output is the robustified DWH test and leads to strong rejection of the null hypothesis that hi_empunion is exogenous. We conclude that it is endogenous.

We obtain exactly the same test statistic when we manually perform the robustified DWH test. We have

```
. * Robust Durbin-Wu-Hausman test of endogeneity implemented manually
. quietly regress hi_empunion ssiratio $x2list

. quietly predict v1hat, resid

. quietly regress ldrugexp hi_empunion v1hat $x2list, vce(robust)

. test v1hat

 ( 1)  v1hat = 0

       F(  1, 10081) =   26.43
            Prob > F =    0.0000
```

4) (Демешев, Борзых, 18.1)

Величины X_i равномерны на отрезке $[-a; 3a]$ и независимы. Есть несколько наблюдений, $X_1 = 0.5, X_2 = 0.7, X_3 = -0.1$.

1. Найдите $\mathbb{E}(X_i)$ и $\mathbb{E}(|X_i|)$.
2. Постройте оценку метода моментов, используя $\mathbb{E}(X_i)$.
3. Постройте оценку метода моментов, используя $\mathbb{E}(|X_i|)$.
4. Постройте оценку обобщёного метода моментов используя моменты $\mathbb{E}(X_i)$, $\mathbb{E}(|X_i|)$ и взвешивающую матрицу.

$$W = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

5. Найдите оптимальную теоретическую взвешивающую матрицу для обобщённого метода моментов
6. Постройте двухшаговую оценку обобщённого метода моментов, начав со взвешивающей матрицы W

Эконометрика, 2017-2018, 1 модуль

Семинары 1 - 2

2.04.18 и 9.04.18 для

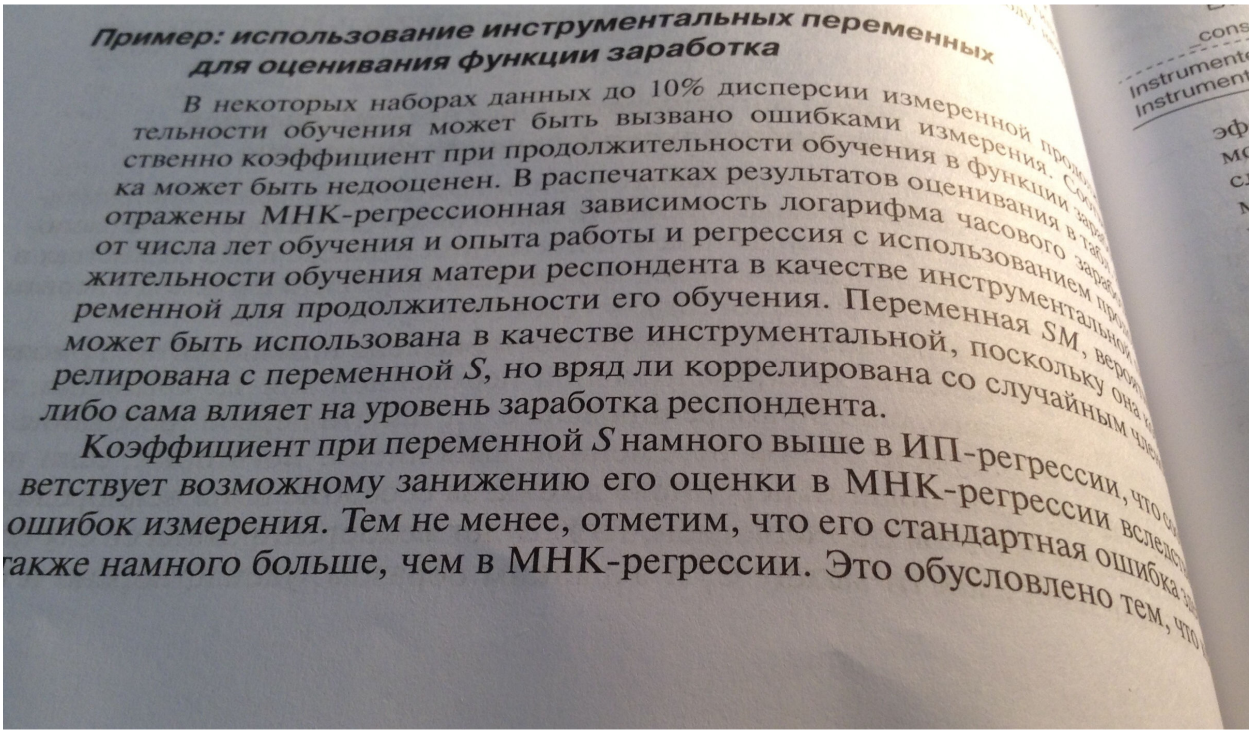
Группы Э_Б2015_Э_3

Семинарист О.А.Демидова

Критика М.Фридменом стандартной функции потребления, раздел 8.5.

1) (Доугерти, 8.7) В некоторой экономике дисперсия переменного дохода составляет 0.5 от дисперсии постоянного дохода, склонность к потреблению товаров кратковременного пользования за счет постоянного дохода составляет 0.6, а расходы на товары длительного пользования отсутствуют. Каким будет значение мультипликатора, полученного на оснве построения «наивной» регрессионной зависимости потребления от дохода, и каково его истинное значение?

2) (Доугерти, раздел 8)



Пример: использование инструментальных переменных для оценивания функции заработка

В некоторых наборах данных до 10% дисперсии измеренной продол-
тельности обучения может быть вызвано ошибками измерения. Со
ственно коэффициент при продолжительности обучения в функции
ка может быть недооценен. В распечатках результатов оценивания в таб
отражены МНК-регрессионная зависимость логарифма часового зара
от числа лет обучения и опыта работы и регрессия с использованием про
жительности обучения матери респондента в качестве инструментальн
ременной для продолжительности его обучения. Переменная *SM*, веро
может быть использована в качестве инструментальной, поскольку она к
релирована с переменной *S*, но вряд ли коррелирована со случайным чле
либо сама влияет на уровень заработка респондента.

Коэффициент при переменной *S* намного выше в ИП-регрессии, что со
ветствует возможному занижению его оценки в МНК-регрессии вследс
ошибок измерения. Тем не менее, отметим, что его стандартная ошибка
также намного больше, чем в МНК-регрессии. Это обусловлено тем, что

тье требо-
ных определи-

(8.52)

ную пере-
льно, с Y_1,
ежду C_1 и
шие по-

ой год,
виатан
чаем-
ырех
оль-
то-
ла
х

Таблица 8.2

reg LGEARN S EXP

Source	SS	df	MS
Model	50.9842581	2	25.492129
Residual	135.723385	537	.252743734
Total	186.707643	539	.34639637

Number of obs = 540
F(2,537) = 100.86
Prob > F = 0.0000
R-squared = 0.2731
Adj R-squared = 0.2704
Root MSE = .50274

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.1235911	.0090989	13.58	0.000	.1057173	.141465
EXP	.0350826	.0050046	7.01	0.000	.0252515	.0449137
_cons	.5093196	.1663823	3.06	0.002	.1824796	.8361596

ivreg LGEARN EXP (S=SM)

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	46.9446075	2	23.4723038
Residual	139.763036	537	.260266361
Total	186.707643	539	.34639637

Number of obs = 540
F(2,537) = 28.38
Prob > F = 0.0000
R-squared = 0.2514
Adj R-squared = 0.2486
Root MSE = .51016

LGEARN	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	.1599676	.0252801	6.33	0.000	.1103076	.2096277
EXP	.0394422	.0058092	6.79	0.000	.0280306	.0508537
_cons	−.0617062	.4061769	−0.15	0.879	−.8595966	.7361841

Instrumented: S

Instruments: EXP SM

…ключ «с…
…бодный член име…
…в данном случае смысл оп…
…поэтому константа включена.

Таблица 8.4

```
. ivreg LGEARN EXP ASVABC MALE ETHBLACK ETHHISP
  (S=SM SF SIBLINGS LIBRARY)
Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =  540
-------------+------------------------------           F(6,533)      =  37.
       Model |  64.4915831     6  10.7485972           Prob > F      =  0.0
    Residual |   122.21606   533  .229298424           R-squared     =  0.3
-------------+------------------------------           Adj R-squared =  0.3
       Total |  186.707643   539   .34639637           Root MSE      =  .47
------------------------------------------------------------------------------
      LGEARN |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
-------------+----------------------------------------------------------------
           S |    .111379   .0476886     2.34   0.020     .0176984    .20
         EXP |   .0258798   .0081187     3.19   0.002     .0099313    .04
      ASVABC |   .0092263    .007991     1.15   0.249    -.0064714    .02
        MALE |   .2619787   .0429283     6.10   0.000     .1776492    .34
    ETHBLACK |  -.0121846   .0822942    -0.15   0.882    -.1738454    .14
     ETHHISP |   .0457639   .0955115     0.48   0.632    -.1418612    .23
       _cons |   .2258512   .3887468     0.58   0.562    -.5378125    .98
------------------------------------------------------------------------------
Instrumented: S
Instruments: EXP ASVABC MALE ETHBLACK ETHHISP SM SF SIBLINGS LIBRARY
------------------------------------------------------------------------------
. estimates store EARNIV
. reg LGEARN S EXP ASVABC MALE ETHBLACK ETHHISP

      Source |       SS       df       MS              Number of obs =  540
-------------+------------------------------           F(6,533)      =  47.9
       Model |   65.490707     6  10.9151178           Prob > F      =  0.00
    Residual |  121.216936   533  .227423895           R-squared     =  0.35
-------------+------------------------------           Adj R-squared =  0.34
       Total |  186.707643   539   .34639637           Root MSE      =  .47
------------------------------------------------------------------------------
      LGEARN |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Inter
-------------+----------------------------------------------------------------
           S |   .0883257   .0109987     8.03   0.000     .0667196    .109
         EXP |   .0227131   .0050095     4.53   0.000     .0128724    .032
      ASVABC |    0129274   .0028834     4.48   0.000     .0072633    .018
        MALE |   .2652878    .042235     6.28   0.000     .1823203    .348
    ETHBLACK |   .0077265   .0715863     0.11   0.914    -.1328994    .148
     ETHHISP |   .0536544   .0937966     0.57   0.568    -.1306019    .237
       _cons |   .4002952   .1663149     2.41   0.016     .0735821    .727
------------------------------------------------------------------------------
```

```
estimates store EARNOLS
hausman EARNIV EARNOLS, constant
```

— Coefficients —

	(b) EARNIV	(B) EARNOLS	(b – B) Difference	sqrt(diag(V_b-V_B)) S.E.
S	.111379	.0883257	.0230533	.0464029
EXP	.0258798	.0227131	.0031667	.0063889
ASVABC	.0092263	.0129274	–.0037011	.0074527
MALE	.2619787	.2652878	–.0033091	.0076842
ETHBLACK	–.0121846	.0077265	–.019911	.0405924
ETHHISP	.0457639	.0536544	–.0078904	.018018
_cons	.2258512	.4002952	–.174444	.3513736

b = consistent under Ho and Ha; obtained from ivreg
B = inconsistent under Ha, efficient under Ho; obtained from regress

Test: Ho: difference in coefficients not systematic

chi2(7) = (b – B)'[(V_b – V_B)^(–1)](b – B) = = 0.25

Prob>chi2 = 0.9999

3) Cameron, Trivedy, Microeconometrics using STATA

5.3.2 Medical expenditures with one endogenous regressor

We consider a model with one endogenous regressor, several exogenous regressors, and one or more excluded exogenous variables that serve as the identifying instruments.

The dataset is an extract from the Medical Expenditure Panel Survey (MEPS) of individuals over the age of 65 years, similar to the dataset described in section 3.2.1. The equation to be estimated has the dependent variable ldrugexp, the log of total out-of-pocket expenditures on prescribed medications. The regressors are an indicator for whether the individual holds either employer or union-sponsored health insurance (hi_empunion), number of chronic conditions (totchr), and four sociodemographic variables: age in years (age), indicators for whether female (female) and whether black or Hispanic (blhisp), and the natural logarithm of annual household income in thousands of dollars (linc).

We treat the health insurance variable hi_empunion as endogenous. The intuitive justification is that having such supplementary insurance on top of the near universal Medicare insurance for the elderly may be a choice variable. Even though most individuals in the sample are no longer working, those who expected high future medical expenses might have been more likely to choose a job when they were working that would provide supplementary health insurance upon retirement. Note that Medicare did not cover drug expenses for the time period we study.

We use the global macro x2list to store the names of the variables that are treated as exogenous regressors. We have

```
. * Read data, define global x2list, and summarize data
. use mus06data.dta

. global x2list totchr age female blhisp linc

. summarize ldrugexp hi_empunion $x2list
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ldrugexp	10391	6.479668	1.363395	0	10.18017
hi_empunion	10391	.3796555	.4853245	0	1
totchr	10391	1.860745	1.290131	0	9
age	10391	75.04639	6.69368	65	91
female	10391	.5797325	.4936256	0	1
blhisp	10391	.1703397	.3759491	0	1
linc	10089	2.743275	.9131433	-6.907755	5.744476

6.3.3 Available instruments

We consider four potential instruments for hi_empunion. Two reflect the income status of the individual and two are based on employer characteristics.

The ssiratio instrument is the ratio of an individual's social security income to the individual's income from all sources, with high values indicating a significant income constraint. The lowincome instrument is a qualitative indicator of low-income status. Both these instruments are likely to be relevant, because they are expected to be negatively correlated with having supplementary insurance. To be valid instruments, we need to assume they can be omitted from the equation for ldrugexp, arguing that the direct role of income is adequately captured by the regressor linc.

The firmsz instrument measures the size of the firm's employed labor force, and the multlc instrument indicates whether the firm is a large operator with multiple locations. These variables are intended to capture whether the individual has access to supplementary insurance through the employer. These two variables are irrelevant for those who are retired, self-employed, or purchase insurance privately. In that sense, these two instruments could potentially be weak.

```
. * Summarize available instruments
. summarize ssiratio lowincome multlc firmsz if linc!=.
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ssiratio	10089	.5365438	.3678175	0	9.25062
lowincome	10089	.1874319	.3902771	0	1
multlc	10089	.0620478	.2412543	0	1
firmsz	10089	.1405293	2.170389	0	50

We have four available instruments for one endogenous regressor. The obvious approach is to use all available instruments, because in theory this leads to the most efficient estimator. In practice, it may lead to larger small-sample bias because the small-sample biases of IV estimators increase with the number of instruments (Hahn and Hausman 2002).

At a minimum, it is informative to use correlate to view the gross correlation between endogenous variables and instruments and between instruments. When multiple instruments are available, as in the case of overidentified models, then it is actually the partial correlation after controlling for other available instruments that matters. This important step is deferred to sections 6.4.2 and 6.4.3.

6.3.4 IV estimation of an exactly identified model

We begin with IV regression of `ldrugexp` on the endogenous regressor `hi_empunion`, instrumented by the single instrument `ssiratio`, and several exogenous regressors.

We use `ivregress` with the `2sls` estimator and the options `vce(robust)` to control for heteroskedastic errors and `first` to provide output that additionally reports results from the first-stage regression. The output is in two parts:

```
. * IV estimation of a just-identified model with single endog regressor
. ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, vce(robust) first

First-stage regressions
-----------------------

                                                  Number of obs   =      10089
                                                  F(   6,  10082) =     119.18
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.0761
                                                  Adj R-squared   =     0.0755
                                                  Root MSE        =     0.4672

------------------------------------------------------------------------------
             |               Robust
 hi_empunion |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      totchr |   .0127865   .0036655     3.49   0.000     .0056015    .0199716
         age |  -.0086323   .0007087   -12.18   0.000    -.0100216   -.0072431
      female |   -.07345    .0096392    -7.62   0.000    -.0923448   -.0545552
      blhisp |   -.06268    .0122742    -5.11   0.000     -.08674    -.0386201
        linc |   .0483937   .0066075     7.32   0.000     .0354417    .0613456
    ssiratio |  -.1916432   .0236326    -8.11   0.000    -.2379678   -.1453186
       _cons |   1.028981   .0581387    17.70   0.000     .9150172    1.142944
------------------------------------------------------------------------------

Instrumental variables (2SLS) regression          Number of obs =      10089
                                                  Wald chi2(6)  =    2000.86
                                                  Prob > chi2   =     0.0000
                                                  R-squared     =     0.0640
                                                  Root MSE      =     1.3177

------------------------------------------------------------------------------
             |               Robust
    ldrugexp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 hi_empunion |  -.8975913   .2211268    -4.06   0.000    -1.330992   -.4641908
      totchr |   .4502655   .0101969    44.16   0.000      .43028     .470251
         age |  -.0132176   .0029977    -4.41   0.000    -.0190931   -.0073421
      female |  -.020406    .0326114    -0.63   0.531    -.0843232    .0435113
      blhisp |  -.2174244   .0394944    -5.51   0.000    -.294832    -.1400167
        linc |   .0870018   .0226356     3.84   0.000     .0426368    .1313668
       _cons |   6.78717    .2688453    25.25   0.000     6.260243    7.314097
------------------------------------------------------------------------------
Instrumented:  hi_empunion
Instruments:   totchr age female blhisp linc ssiratio
```

6.3.6 Testing for regressor endogeneity

The preceding analysis treats the insurance variable, `hi_empunion`, as endogenous. If instead the variable is exogenous, then the IV estimators (IV, 2SLS, or GMM) are still consistent, but they can be much less efficient than the OLS estimator.

The Hausman test principle provides a way to test whether a regressor is endogenous. If there is little difference between OLS and IV estimators, then there is no need to instrument, and we conclude that the regressor was exogenous. If instead there is considerable difference, then we needed to instrument and the regressor is endogenous. The test usually compares just the coefficients of the endogenous variables. In the case of just one potentially endogenous regressor with a coefficient denoted by β, the Hausman test statistic

$$T_H = \frac{(\widehat{\beta}_{\text{IV}} - \widehat{\beta}_{\text{OLS}})^2}{\widehat{V}(\widehat{\beta}_{\text{IV}} - \widehat{\beta}_{\text{OLS}})}$$

is $\chi^2(1)$ distributed under the null hypothesis that the regressor is exogenous.

Before considering implementation of the test, we first obtain the OLS estimates to compare them with the earlier IV estimates. We have

```
. * Obtain OLS estimates to compare with preceding IV estimates
. regress ldrugexp hi_empunion $x2list, vce(robust)

Linear regression                                   Number of obs =    10089
                                                    F(  6, 10082)    876.85
                                                    Prob > F         0.0000
                                                    R-squared        0.1770
                                                    Root MSE        =1.236

------------------------------------------------------------------------------
             |               Robust
    ldrugexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 hi_empunion |   .0738788   .0259848     2.84   0.004      .0229435    .1248141
      totchr |   .4403807   .0093633    47.03   0.000      .4220268    .4587346
         age |  -.0035295    .001937    -1.82   0.068     -.0073264    .0002675
      female |   .0578055   .0253651     2.28   0.023      .0080848    .1075262
      blhisp |  -.1513068   .0341264    -4.43   0.000     -.2182013   -.0844122
        linc |   .0104815   .0137126     0.76   0.445     -.0163979     .037361
       _cons |   5.861131   .1571037    37.31   0.000      5.553176    6.169085
------------------------------------------------------------------------------
```

The OLS estimates differ substantially from the just-identified IV estimates given in section 6.3.4. The coefficient of `hi_empunion` has an OLS estimate of 0.074, greatly different from the IV estimate of −0.898. This is strong evidence that `hi_empunion` is endogenous. Some coefficients of exogenous variables also change, notably, those for age and `female`. Note also the loss in precision in using IV. Most notably, the standard error of the instrumented regressor increases from 0.026 for OLS to 0.221 for IV, an eightfold increase, indicating the potential loss in efficiency due to IV estimation.

The `hausman` command can be used to compute T_H under the assumption that $\widehat{V}(\widehat{\beta}_{\mathrm{IV}} - \widehat{\beta}_{\mathrm{OLS}}) = \widehat{V}(\widehat{\beta}_{\mathrm{IV}}) - \widehat{V}(\widehat{\beta}_{\mathrm{OLS}})$; see section 12.7.5. This greatly simplifies analysis because then all that is needed are coefficient estimates and standard errors from separate IV estimation (IV, 2SLS, or GMM) and OLS estimation. But this assumption is too strong. It is correct only if $\widehat{\beta}_{\mathrm{OLS}}$ is the fully efficient estimator under the null hypothesis of exogeneity, an assumption that is valid only under the very strong assumption that model errors are independent and homoskedastic. One possible variation is to perform an appropriate bootstrap; see section 13.4.6.

The postestimation `estat endogenous` command implements the related Durbin–Wu–Hausman (DWH) test. Because the DWH test uses the device of augmented regressors, it produces a robust test statistic (Davidson 2000). The essential idea is the following. Consider the model as specified in section 6.2.1. Rewrite the structural equation (6.2) with an additional variable, v_1, that is the error from the first-stage equation (6.3) for y_2. Then

$$y_{1i} = \beta_1 y_{2i} + \mathbf{x}_{1i}'\beta_2 + \rho v_{1i} + u_i$$

Under the null hypothesis that y_2 is exogenous, $E(v_{1i}u_i|y_{2i}, \mathbf{x}_{1i}) = 0$. If v_1 could be observed, then the test of exogeneity would be the test of $H_0: \rho = 0$ in the OLS regression of y_1 on y_2, $\mathbf{x}_1$, and v_1. Because v_1 is not directly observed, the fitted residual vector

$\hat{v}_1$ from the first-stage OLS regression (6.3) is instead substituted. For independent homoskedastic errors, this test is asymptotically equivalent to the earlier Hausman test. In the more realistic case of heteroskedastic errors, the test of $H_0 : \rho = 0$ can still be implemented provided that we use robust variance estimates. This test can be extended to the multiple endogenous regressors case by including multiple residual vectors and testing separately for correlation of each with the error on the structural equation.

We apply the test to our example with one potentially endogenous regressor, hi_empunion, instrumented by ssiratio. Then

```
. * Robust Durbin-Wu-Hausman test of endogeneity implemented by estat endogenous
. ivregress 2sls ldrugexp (hi_empunion = ssiratio) $x2list, vce(robust)

Instrumental variables (2SLS) regression               Number of obs =   10089
                                                       Wald chi2(6)  = 2000.86
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.0640
                                                       Root MSE      =  1.3177

------------------------------------------------------------------------------
             |               Robust
    ldrugexp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 hi_empunion |  -.8975913   .2211268    -4.06   0.000    -1.330992   -.4641908
      totchr |   .4502655   .0101969    44.16   0.000      .43028     .470251
         age |  -.0132176   .0029977    -4.41   0.000    -.0190931   -.0073421
      female |   -.020406   .0326114    -0.63   0.531    -.0843232    .0435113
      blhisp |  -.2174244   .0394944    -5.51   0.000     -.294832   -.1400167
       linc |   .0870018   .0226356     3.84   0.000     .0426368    .1313668
       _cons |    6.78717   .2688453    25.25   0.000     6.260243    7.314097
------------------------------------------------------------------------------
Instrumented:  hi_empunion
Instruments:   totchr age female blhisp linc ssiratio

. estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Robust score chi2(1)            =  24.935  (p = 0.0000)
  Robust regression F(1,10081)    = 26.4333  (p = 0.0000)
```

The last line of output is the robustified DWH test and leads to strong rejection of the null hypothesis that hi_empunion is exogenous. We conclude that it is endogenous.

We obtain exactly the same test statistic when we manually perform the robustified DWH test. We have

```
. * Robust Durbin-Wu-Hausman test of endogeneity implemented manually
. quietly regress hi_empunion ssiratio $x2list

. quietly predict v1hat, resid

. quietly regress ldrugexp hi_empunion v1hat $x2list, vce(robust)

. test v1hat

 ( 1)  v1hat = 0

       F(  1, 10081) =   26.43
            Prob > F =    0.0000
```

3.7 Tests of overidentifying restrictions

The validity of an instrument cannot be tested in a just-identified model. But it is possible to test the validity of overidentifying instruments in an overidentified model provided that the parameters of the model are estimated using optimal GMM. The same test has several names, including overidentifying restrictions (OIR) test, overidentified (OID) test, Hansen's test, Sargan's test, and Hansen–Sargan test.

The starting point is the fitted value of the criterion function (6.8) after optimal GMM, i.e., $Q(\widehat{\beta}) = \{(1/N)(\mathbf{y} - \mathbf{X}\widehat{\beta})'\mathbf{Z}\}\widehat{\mathbf{S}}^{-1}\{(1/N)\mathbf{Z}'(\mathbf{y} - \mathbf{X}\widehat{\beta})\}$. If the population moment conditions $E\{\mathbf{Z}'(\mathbf{y} - \mathbf{X}\beta)\} = \mathbf{0}$ are correct, then $\mathbf{Z}'(\mathbf{y} - \mathbf{X}\widehat{\beta}) \simeq \mathbf{0}$, so $Q(\widehat{\beta})$ should be close to zero. Under the null hypothesis that all instruments are valid, it can be shown that $Q(\widehat{\beta})$ has an asymptotic chi-squared distribution with degrees of freedom equal to the number of overidentifying restrictions.

Large values of $Q(\widehat{\beta})$ lead to rejection of H_0: $E\{\mathbf{Z}'(\mathbf{y} - \mathbf{X}\beta)\} = \mathbf{0}$. Rejection is interpreted as indicating that at least one of the instruments is not valid. Tests can have power in other directions, however, as emphasized in section 3.5.5. It is possible that rejection of H_0 indicates that the model $\mathbf{X}\beta$ for the conditional mean is misspecified. Going the other way, the test is only one of validity of the overidentifying instruments, so failure to reject H_0 does not guarantee that all the instruments are valid.

The test is implemented with the postestimation `estat overid` command following the `ivregress gmm` command for an overidentified model. We do so for the optimal GMM estimator with heteroskedastic errors and instruments, `ssiratio` and `multc`. The example below implements `estat overid` under the overidentifying restriction.

```
. * Test of overidentifying restrictions following ivregress gmm
. quietly ivregress gmm ldrugexp (hi_empunion = ssiratio multlc)
> $x2list, wmatrix(robust)
. estat overid
  Test of overidentifying restriction:
  Hansen's J chi2(1) = 1.04754 (p = 0.3061)
```

The test statistic is $\chi^2(1)$ distributed because the number of overidentifying restrictions equals $2 - 1 = 1$. Because $p > 0.05$, we do not reject the null hypothesis and conclude that the overidentifying restriction is valid.

A similar test using all four available instruments yields

```
. * Test of overidentifying restrictions following ivregress gmm
. ivregress gmm ldrugexp (hi_empunion = ssiratio lowincome multlc firmsz)
> $x2list, wmatrix(robust)
Instrumental variables (GMM) regression               Number of obs =    10089
                                                      Wald chi2(6)  = 2042.12
                                                      Prob > chi2   =  0.0000
                                                      R-squared     =  0.0829
GMM weight matrix: Robust                             Root MSE      =  1.3043
```

ldrugexp	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
hi_empunion	-.8124043	.1846433	-4.40	0.000	-1.174299	-.45051
totchr	.449488	.010047	44.74	0.000	.4297962	.4691799
age	-.0124598	.0027466	-4.54	0.000	-.0178432	-.0070765
female	-.0104528	.0306889	-0.34	0.733	-.0706019	.0496963
blhisp	-.2061018	.0382891	-5.38	0.000	-.2811471	-.1310566
linc	.0796532	.0203397	3.92	0.000	.0397882	.1195183
_cons	6.7126	.2425973	27.67	0.000	6.237118	7.188081

```
Instrumented:  hi_empunion
Instruments:   totchr age female blhisp linc ssiratio lowincome multlc
               firmsz
. estat overid
  Test of overidentifying restriction:
  Hansen's J chi2(3) = 11.5903 (p = 0.0089)
```

Now we reject the null hypothesis at level 0.05 and, barely, at level 0.01. Despite this rejection, the coefficient of the endogenous regressor hi_empunion is −0.812, not all that different from the estimate when ssiratio is the only instrument.

4) (Демешев, Борзых, 18.1)

Величины X_i равномерны на отрезке $[-a; 3a]$ и независимы. Есть несколько наблюдений, $X_1 = 0.5, X_2 = 0.7, X_3 = -0.1$.

1. Найдите $\mathbb{E}(X_i)$ и $\mathbb{E}(|X_i|)$.
2. Постройте оценку метода моментов, используя $\mathbb{E}(X_i)$.
3. Постройте оценку метода моментов, используя $\mathbb{E}(|X_i|)$.
4. Постройте оценку обобщёного метода моментов используя моменты $\mathbb{E}(X_i)$, $\mathbb{E}(|X_i|)$ и взвешивающую матрицу.

$$W = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

5. Найдите оптимальную теоретическую взвешивающую матрицу для обобщённого метода моментов
6. Постройте двухшаговую оценку обобщённого метода моментов, начав со взвешивающей матрицы W