linearmodels 4.5 (../../index.html)

TABLE OF CONTENTS

Contents:

Search...

Three-stage Least Squares (3SLS)

This example demonstrates how a system of simultaneous equations can be jointly estimated using three-stage least squares (3SLS). The simultaneous equations model the wage and number of hours worked. The two equations are

$$hours = \beta_0 + \beta_1 \ln(wage) + \beta_2 educ + \beta_3 age + \beta_4 kidslt6 + \beta_5 nwifeinc + \epsilon_i^h$$
$$\ln(wage) = \gamma_0 + \gamma_1 hours + \gamma_2 educ + \gamma_3 educ^2 + \gamma_4 exper + \epsilon_i^w$$

Each equation has a single exogenous variables. The instruments for the endogenous variables are the regressors that appear in one equation but not the other.

Data ¶

The data set is the MORZ data set from Wooldridge (2002).

In [1]:

```
from linearmodels.datasets import mroz
data = mroz.load()
```

Here the relevant variables are selected and missing observations are dropped to avoid warnings.

In [2]:

```
data = data[['hours','educ','age','kidslt6','nwifeinc','lwage','exper',
data = data.dropna()
```

The main models are imported:

- `IV2SLS` - single equation 2-stage least squares
- `IV3SLS` - system estimation using instrumental variables
- `SUR` - system estimation without endogenous variables

In [3]:

```
from linearmodels import IV2SLS, IV3SLS, SUR, IVSystemGMM
```

Formulas

These examples use the formula interface. This is usually simpler when models have exogenous regressors, endogenous regressors and instruments. The syntax is the same as in the 2SLS models.

In [4]:

```
hours = 'hours ~ educ + age + kidslt6 + nwifeinc + [lwage ~ exper + exp

hours_mod = IV2SLS.from_formula(hours, data)
hours_res = hours_mod.fit(cov_type='unadjusted')
print(hours_res)
```

```
                          IV-2SLS Estimation Summary
======================================================================
Dep. Variable:                  hours   R-squared:
Estimator:                    IV-2SLS   Adj. R-squared:
No. Observations:                 428   F-statistic:
Date:                Mon, Dec 11 2017   P-value (F-stat)
Time:                        20:42:20   Distribution:
Cov. Estimator:            unadjusted

                             Parameter Estimates
======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
----------------------------------------------------------------------
educ          -99.299     48.997    -2.0266     0.0427     -195.33
age            19.429     6.2770     3.0952     0.0020      7.1261
kidslt6       -51.616     183.96    -0.2806     0.7790     -412.18
nwifeinc      -11.445     6.6787    -1.7137     0.0866     -24.535
lwage          1626.3     472.19     3.4442     0.0006      700.85
======================================================================

Endogenous: lwage
Instruments: exper, expersq
Unadjusted Covariance (Homoskedastic)
Debiased: False
```

The $\ln$ wage model can be similarly specified and estimated

In [5]:

```
lwage = 'lwage ~ educ + exper + expersq + [hours ~ age + kidslt6 + nwif

lwage_mod = IV2SLS.from_formula(lwage, data)
lwage_res = lwage_mod.fit(cov_type='unadjusted')
print(lwage_res)
```

```
                          IV-2SLS Estimation Summary
======================================================================
Dep. Variable:                  lwage   R-squared:
Estimator:                    IV-2SLS   Adj. R-squared:
No. Observations:                 428   F-statistic:
Date:                Mon, Dec 11 2017   P-value (F-stat)
Time:                        20:42:20   Distribution:
Cov. Estimator:            unadjusted

                             Parameter Estimates
======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
----------------------------------------------------------------------
educ           0.0875     0.0162     5.3892     0.0000      0.0557
exper          0.0524     0.0299     1.7501     0.0801     -0.0063
expersq       -0.0009     0.0006    -1.4898     0.1363     -0.0021
hours         -0.0003     0.0003    -0.8666     0.3862     -0.0009
======================================================================

Endogenous: hours
Instruments: age, kidslt6, nwifeinc
Unadjusted Covariance (Homoskedastic)
Debiased: False
```

A system can be specified using a dictionary for formulas. The dictionary keys are used as equation labels. Aside from this simple change, the syntax is identical.

Here the model is estimated using `method='ols'` which will just simultaneously estimate the two equations but will produce estimates that are identical to separate equations.

In [6]:

```
equations = dict(hours=hours, lwage=lwage)
system_2sls = IV3SLS.from_formula(equations, data)
system_2sls_res = system_2sls.fit(method='ols', cov_type='unadjusted')
print(system_2sls_res)
```

linearmodels 4.5 (../../index.html)

TABLE OF CONTENTS

Contents:

Search...

```
                        System OLS Estimation Summary
======================================================================
Estimator:                        OLS   Overall R-squared:
No. Equations.:                     2   Cov. Estimator:            una
No. Observations:                 428   Num. Constraints:
Date:                Mon, Dec 11 2017
Time:                        20:42:20


                  Equation: hours, Dependent Variable: hours
======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
----------------------------------------------------------------------
educ          -99.299     48.997    -2.0266     0.0427     -195.33
age            19.429     6.2770     3.0952     0.0020      7.1261
kidslt6       -51.616     183.96    -0.2806     0.7790     -412.18
nwifeinc      -11.445     6.6787    -1.7137     0.0866     -24.535
lwage          1626.3     472.19     3.4442     0.0006      700.85

==============
 Instruments
--------------
exper, expersq

                  Equation: lwage, Dependent Variable: lwage
======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
----------------------------------------------------------------------
educ           0.0875     0.0162     5.3892     0.0000      0.0557
exper          0.0524     0.0299     1.7501     0.0801     -0.0063
expersq       -0.0009     0.0006    -1.4898     0.1363     -0.0021
hours         -0.0003     0.0003    -0.8666     0.3862     -0.0009
======================
     Instruments
----------------------
age, kidslt6, nwifeinc
----------------------

Covariance Estimator:
Homoskedastic (Unadjusted) Covariance (Debiased: False, GLS: False)
```

Using `method='gls'` will use GLS estimates which can be more efficient than the usual estimates. Here only the first equation changes. This is due to the structure of the problem.

In [7]:

```
equations = dict(hours=hours, lwage=lwage)
system_3sls = IV3SLS.from_formula(equations, data)
system_3sls_res = system_3sls.fit(method='gls', cov_type='unadjusted')
print(system_3sls_res)
```

```
                          System GLS Estimation Summary
=======================================================================
Estimator:                        GLS   Overall R-squared:
No. Equations.:                     2   Cov. Estimator:             una
No. Observations:                 428   Num. Constraints:
Date:                Mon, Dec 11 2017
Time:                        20:42:20


                  Equation: hours, Dependent Variable: hours
=======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
-----------------------------------------------------------------------
educ          -109.90     48.052    -2.2870     0.0222     -204.08
age            13.651     5.5456     2.4617     0.0138      2.7822
kidslt6       -196.61     170.04    -1.1563     0.2476     -529.88
nwifeinc      -6.4136     5.4646    -1.1736     0.2405     -17.124
lwage          1872.7     461.13     4.0611     0.0000      968.88

==============
 Instruments
--------------
exper, expersq

                  Equation: lwage, Dependent Variable: lwage
=======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
-----------------------------------------------------------------------
educ           0.0859     0.0159     5.3864     0.0000      0.0546
exper          0.0550     0.0295     1.8622     0.0626     -0.0029
expersq       -0.0010     0.0006    -1.7539     0.0794     -0.0022
hours         -0.0003     0.0003    -0.8398     0.4010     -0.0009
======================
     Instruments
----------------------
age, kidslt6, nwifeinc
----------------------

Covariance Estimator:
Homoskedastic (Unadjusted) Covariance (Debiased: False, GLS: True)
```

Direct Model Specification

The model can be directly specified using a dictionary of dictionaries where the inner dictionaries contain the 4 components of the model:

- dependent - The dependent variable
- exog - Exogenous regressors
- endog - Endogenous regressors
- instruments - Instrumental variables

The estimates are the same. This interface is more useful for programatically generating and estimating models.

In [8]:

```
hours = {'dependent': data[['hours']],
         'exog': data[['educ','age','kidslt6','nwifeinc']],
         'endog': data[['lwage']],
         'instruments': data[['exper','expersq']]}


lwage = {'dependent': data[['lwage']],
         'exog': data[['educ','exper','expersq']],
         'endog': data[['hours']],
         'instruments': data[['age','kidslt6','nwifeinc']]}

equations = dict(hours=hours, lwage=lwage)
system_3sls = IV3SLS(equations)
system_3sls_res = system_3sls.fit(cov_type='unadjusted')
print(system_3sls_res)
```

```
                           System GLS Estimation Summary
=======================================================================
Estimator:                        GLS   Overall R-squared:
No. Equations.:                     2   Cov. Estimator:            una
No. Observations:                 428   Num. Constraints:
Date:                Mon, Dec 11 2017
Time:                        20:42:21


                 Equation: hours, Dependent Variable: hours
=======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
-----------------------------------------------------------------------
educ          -109.90     48.052    -2.2870     0.0222     -204.08
age            13.651     5.5456     2.4617     0.0138      2.7822
kidslt6       -196.61     170.04    -1.1563     0.2476     -529.88
nwifeinc      -6.4136     5.4646    -1.1736     0.2405     -17.124
lwage          1872.7     461.13     4.0611     0.0000      968.88

==============
 Instruments
--------------
exper, expersq

                 Equation: lwage, Dependent Variable: lwage
=======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
-----------------------------------------------------------------------
educ           0.0859     0.0159     5.3864     0.0000      0.0546
exper          0.0550     0.0295     1.8622     0.0626     -0.0029
expersq       -0.0010     0.0006    -1.7539     0.0794     -0.0022
hours         -0.0003     0.0003    -0.8398     0.4010     -0.0009
======================
     Instruments
----------------------
age, kidslt6, nwifeinc
----------------------

Covariance Estimator:
Homoskedastic (Unadjusted) Covariance (Debiased: False, GLS: True)
```

System GMM Estimation

System GMM is an alternative to 3SLS estimation. It is the natural extension to GMM estimation of IV models. It makes weaker assumptions about instruments than 3SLS does. In particular, instruments are assumed exogenous on an equation-by-equation basis rather than the 3SLS assumption that all instruments are exogenous in all equations.

The system GMM estimator is similar to the 3SLS estimator except that it requires making a choice about the moment weighting estimator. Valid options for the weighting estimator are `'unadjusted'` or `'homoskedastic'` which assumes that residuals are conditionally homoskedastic or `'robust'` or `'heteroskedastic'` which allows for conditional heteroskedasticity.

The System GMM estimator also supports iterative application where it is possible to iterate until convergence.

Here the examples make use of the same data as in the 3SLS example and only use the formula interface. The default uses 2-step (efficient) GMM.

In [9]:

```
equations = dict(hours='hours ~ educ + age + kidslt6 + nwifeinc + [lwag
                 lwage='lwage ~ educ + exper + expersq + [hours ~ age +
system_gmm = IVSystemGMM.from_formula(equations, data, weight_type='una
system_gmm_res = system_gmm.fit(cov_type='unadjusted')
print(system_gmm_res)
```

Search...

```
                   System 2-Step System GMM Estimation Summary
======================================================================
Estimator:          2-Step System GMM   Overall R-squared:
No. Equations.:                     2   Cov. Estimator:            una
No. Observations:                 428   Num. Constraints:
Date:                Mon, Dec 11 2017
Time:                        20:42:21


                    Equation: hours, Dependent Variable: hours
======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
----------------------------------------------------------------------
educ          -109.89     48.038    -2.2876     0.0222     -204.05
age            13.653     5.5440     2.4626     0.0138      2.7868
kidslt6       -196.57     169.99    -1.1564     0.2475     -529.74
nwifeinc      -6.4149     5.4631    -1.1742     0.2403     -17.122
lwage          1872.6     461.00     4.0621     0.0000      969.08

==============
 Instruments
--------------
exper, expersq

                    Equation: lwage, Dependent Variable: lwage
======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
----------------------------------------------------------------------
educ           0.0859     0.0159     5.3866     0.0000      0.0546
exper          0.0550     0.0295     1.8623     0.0626     -0.0029
expersq       -0.0010     0.0006    -1.7540     0.0794     -0.0022
hours         -0.0003     0.0003    -0.8398     0.4010     -0.0009
======================
     Instruments
----------------------
age, kidslt6, nwifeinc
----------------------

Covariance Estimator:
GMM Homoskedastic (Unadjusted) Covariance
Weight Estimator:
Homoskedastic (Unadjusted) Weighting (Debiased: False, Center: False)
```

Robust weighting can be used by setting the `weight_type`. The number of iterations can be set using `iter_limit`. Overall the parameters do not meaningfully change.

In [10]:

```
system_gmm = IVSystemGMM.from_formula(equations, data, weight_type='rob
system_gmm_res = system_gmm.fit(cov_type='robust', iter_limit=100)
print('Number of iterations: ' + str(system_gmm_res.iterations))
print(system_gmm_res)
```

linearmodels 4.5 (../../index.html)

TABLE OF CONTENTS

Contents:

Search...

```
Number of iterations: 20
                 System Iterative System GMM Estimation Summary
======================================================================
Estimator:         Iterative System GMM   Overall R-squared:
No. Equations.:                       2   Cov. Estimator:
No. Observations:                   428   Num. Constraints:
Date:                  Mon, Dec 11 2017
Time:                          20:42:21


                  Equation: hours, Dependent Variable: hours
======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
----------------------------------------------------------------------
educ          -118.31     57.508    -2.0572     0.0397     -231.02
age            15.027     5.9903     2.5086     0.0121      3.2864
kidslt6       -227.22     216.54    -1.0493     0.2940     -651.64
nwifeinc      -8.2578     5.3146    -1.5538     0.1202     -18.674
lwage          1933.5     592.77     3.2618     0.0011      771.69

==============
 Instruments
--------------
exper, expersq

                  Equation: lwage, Dependent Variable: lwage
======================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    U
----------------------------------------------------------------------
educ           0.0967     0.0190     5.0999     0.0000      0.0595
exper          0.0720     0.0363     1.9840     0.0473      0.0009
expersq       -0.0013     0.0007    -1.8475     0.0647     -0.0028   8.
hours         -0.0005     0.0004    -1.2298     0.2188     -0.0013
======================
     Instruments
----------------------
age, kidslt6, nwifeinc
----------------------

Covariance Estimator:
GMM Heteroskedastic (Robust) Covariance
Weight Estimator:
Heteroskedastic (Robust) Weighting (Debiased: False, Center: False)
```

Using formulas to specify models (formulas.html)

Module Reference (../reference.html)