ardl: Stata module to estimate autoregressive distributed lag models

Sebastian Kripfganz[1] Daniel C. Schneider[2]

[1]University of Exeter Business School, Department of Economics, Exeter, UK

[2]Max Planck Institute for Demographic Research, Rostock, Germany

Stata Conference

Chicago, July 29, 2016

```
net install ardl, from(http://www.kripfganz.de/stata/)
```

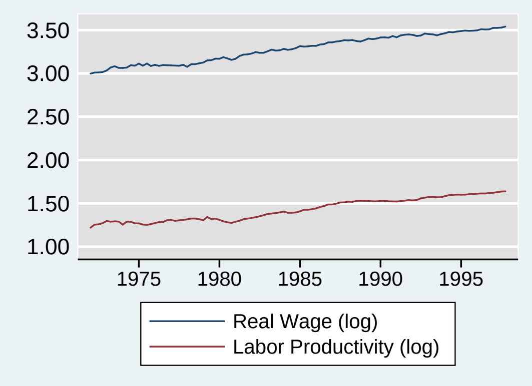
ARDL: autoregressive distributed lag model

- The autoregressive distributed lag (ARDL)[1] model is being used for decades to model the relationship between (economic) variables in a single-equation time-series setup.
- Its popularity also stems from the fact that cointegration of nonstationary variables is equivalent to an error-correction (EC) process, and the ARDL model has a reparameterization in EC form (Engle and Granger, 1987; Hassler and Wolters, 2006).
- The existence of a long-run / cointegrating relationship can be tested based on the EC representation. A bounds testing procedure is available to draw conclusive inference without knowing whether the variables are integrated of order zero or one, $I(0)$ or $I(1)$, respectively (Pesaran, Shin, and Smith, 2001).

[1] Another commonly used abbreviation is ADL.

ARDL: autoregressive distributed lag model

- Long-run relationship: Some time series are bound together due to equilibrium forces even though the individual time series might move considerably.



Data source: Pesaran, Shin, and Smith (2001).

ARDL: autoregressive distributed lag model

- The first public version of the `ardl` command for the estimation of ARDL / EC models and the bounds testing procedure in Stata has been released on August 4, 2014.
- Some indications for the popularity of the ARDL model:
  - Google Scholar returns about 13,200 results when searching for "autoregressive distributed lag", and more than 5,200 citations for the bounds testing paper by Pesaran, Shin, and Smith (2001).
  - A sequence of blog posts by David Giles on ARDL model estimation attracted more than 500 comments.
  - The discussion topic on the `ardl` command is ranked second on Statalist in terms of replies (>100) and views (>20,000).[2]
  - There are already at least 2 independent video tutorials available on the web dealing with the `ardl` command for Stata.

[2] www.statalist.org/forums/forum/general-stata-discussion/general/95329-ardl-in-stata

Estimating long-run relationships

- Engle and Granger (1987) two-step approach for testing the existence of a long-run relationship:
  - Assumption: $(y_t, \mathbf{x}_t)'$ is a vector of $I(1)$ variables.

1. Run an OLS regression for the model in levels:

$$y_t = b_0 + \boldsymbol{\theta}'\mathbf{x}_t + v_t,$$

   and test whether the residuals $\hat{v}_t = y_t - \hat{b}_0 - \hat{\boldsymbol{\theta}}'\mathbf{x}_t$ are stationary (e.g. with a Dickey-Fuller test).

2. Estimate an EC model with the lagged residuals from the first step included as EC term (provided they are stationary):

$$\Delta y_t = c_0 + \gamma \hat{v}_{t-1} + \sum_{i=1}^{p-1} \psi_{yi} \Delta y_{t-i} + \sum_{i=0}^{q-1} \boldsymbol{\psi}_{xi}' \Delta \mathbf{x}_{t-i} + u_t,$$

   and test whether $-1 \leq \gamma < 0$.

- Stata module `egranger` by Mark E. Schaffer (2010) on SSC.

Estimating long-run relationships

- Disadvantages of the Engle and Granger (1987) approach:
  - The order of integration of the variables needs to be determined first.
  - OLS estimation of the static levels model may create bias in finite samples due to the omitted short-run dynamics (Banerjee, Dolado, Hendry, and Smith, 1986).
  - The bias from the first step transmits to poor second-step estimates.
  - The asymptotic distribution of the OLS estimator for the long-run parameters $\boldsymbol{\theta}$ is non-normal, invalidating standard inference based on the t-statistic.
  - General pretesting problems: misclassification of variables as $I(0)$ or $I(1)$; false positives and false negatives at the first step.
- Phillips and Hansen (1990) proposed the fully-modified OLS estimator to overcome some of these problems.

Estimating long-run relationships

- Pesaran and Shin (1998) suggest to obtain the long-run parameters from an ARDL model:
  - OLS estimators of the short-run parameters are $\sqrt{T}$-consistent and asymptotically normal.
  - The corresponding estimators of the long-run parameters are super-consistent if the regressors are $I(1)$, and asymptotically normally distributed irrespective of the order of integration.
- Bounds procedure for testing the existence of a long-run relationship based on the EC representation of the ARDL model:
  - Pesaran, Shin, and Smith (2001) tabulate asymptotic critical values that span a band from all regressors being purely $I(0)$ to all regressors being purely $I(1)$.
  - Narayan (2005) computes corresponding small-sample critical values for various sample sizes.

ARDL model

- ARDL$(p, q, \dots, q)$ model:

$$y_t = c_0 + c_1 t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=0}^{q} \beta_i' \mathbf{x}_{t-i} + u_t,$$

$t = \max(p, q), \dots, T$, for simplicity assuming that the lag order q is the same for all variables in the $K \times 1$ vector $\mathbf{x}_t$.

- The variables in $(y_t, \mathbf{x}_t')'$ are allowed to be purely $I(0)$, purely $I(1)$, or cointegrated.[3]
- The optimal lag orders p and q (possibly different across regressors) can be obtained my minimizing a model selection criterion, e.g. the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).[4]

[3] For a full set of assumptions see Pesaran, Shin, and Smith (2001).

[4] The BIC is also known as the Schwarz or Schwarz-Bayesian information criterion.

EC representation

- Reparameterization in conditional EC form:

$$\Delta y_t = c_0 + c_1 t - \alpha(y_{t-1} - \boldsymbol{\theta}\mathbf{x}_{t-1}) + \sum_{i=1}^{p-1} \psi_{yi} \Delta y_{t-i} + \boldsymbol{\omega}' \Delta \mathbf{x}_t + \sum_{i=1}^{q-1} \boldsymbol{\psi}'_{xi} \Delta \mathbf{x}_{t-i} + u_t,$$

with the speed-of-adjustment coefficient $\alpha = 1 - \sum_{j=1}^{p} \phi_i$ and the long-run coefficients $\boldsymbol{\theta} = \frac{\sum_{j=0}^{q} \boldsymbol{\beta}_j}{\alpha}$.

- Alternative EC parameterization:

$$\Delta y_t = c_0 + c_1 t - \alpha(y_{t-1} - \boldsymbol{\theta}\mathbf{x}_t) + \sum_{i=1}^{p-1} \psi_{yi} \Delta y_{t-i} + \sum_{i=0}^{q-1} \boldsymbol{\psi}'_{xi} \Delta \mathbf{x}_{t-i} + u_t.$$

Testing the existence of a long-run relationship

- Pesaran, Shin, and Smith (2001) approach:
  1. Decide about the inclusion of deterministic model components and obtain the optimal lag orders p and q based on a suitable model selection criterion, e.g. AIC or BIC. (When in doubt, choose higher lag orders for testing purposes.)
  2. Estimate the chosen ARDL$(p, q, \dots, q)$ model by OLS.
  3. Compute the F-statistic for the joint null hypothesis $H_0^F : (\alpha = 0) \cap \left(\sum_{j=0}^{q} \boldsymbol{\beta}_j = \mathbf{0}\right)$ and compare it to the critical values.
  4. If H_0^F is rejected, compute the t-statistic for the single null hypothesis $H_0^t : \alpha = 0$ and compare it to the critical values.
  5. Potentially re-estimate a parsimonious version of the ARDL / EC model.

Testing the existence of a long-run relationship

- Pesaran, Shin, and Smith (2001) provide lower and upper bounds for the asymptotic critical values depending on the number of regressors, their order of integration, and the deterministic model components:
  1. No intercept, no time trend.
  2. Restricted intercept, no time trend.
  3. Unrestricted intercept, no time trend.
  4. Unrestricted intercept, restricted time trend.
  5. Unrestricted intercept, unrestricted time trend.
- Test decisions:
  - Do not reject H_0^F or H_0^t, respectively, if the test statistic is closer to zero than the lower bound of the critical values.
  - Reject the H_0^F or H_0^t, respectively, if the test statistic is more extreme than the upper bound of the critical values.
- The existence of a (conditional) long-run relationship is confirmed if both H_0^F and H_0^t are rejected.

Stata syntax of the ardl command

- Syntax:
  ardl depvar [indepvars] [if] [in] [, options]
- Selected options:
  - lags(numlist): set lag lengths,
  - maxlags(numlist): set maximum lag lengths,
  - ec: display output in error-correction form,
  - ec1: like option ec, but level variables in $t-1$ instead of t,
  - aic: use AIC as information criterion instead of BIC,
  - exog(varlist): exogenous variables in the regression,
  - noconstant: suppress constant term,
  - trendvar(varname): specify trend variable,
  - restricted: restrict constant or trend term.
- Postestimation commands:
  - estat btest: bounds test,
  - predict: fitted values, residuals, and error-correction term,
  - estat ic, nlcom, test, ...

Example

- Pesaran, Shin, and Smith (2001) estimate a UK earnings equation. We focus on the model with unrestricted intercept and no time trend (case 3).

```
. describe w prod ur wedge union d*

              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
w               double  %6.2fc                Real Wage (log)
prod            double  %6.2fc                Labor Productivity (log)
ur              double  %6.2fc                Unemployment Rate (log)
wedge           double  %7.3fc                Wedge (log)
union           double  %7.3fc                Union Power (log)
d7475           byte    %3.0fc                Dummy: Years 1974-1975
d7579           byte    %3.0fc                Dummy: Years 1975-1979

. summarize w prod ur wedge union d*, separator(7)

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
           w |        116    3.252843    .1792393    2.91503    3.54119
        prod |        116    1.402628    .1411384   1.155336    1.63873
          ur |        112    1.838737    .6734774   .0262613    2.53305
       wedge |        116   -.3412168    .0402661  -.4151654  -.2330282
       union |        116   -.6886907    .0602385  -.8461587   -.650586
       d7475 |        116    .0689655    .2544948          0          1
       d7579 |        116    .1724138    .3793785          0          1
```

Example: ARDL model with optimal lag orders

```
. ardl w prod ur wedge union if tin(1972q1,1997q4), exog(d7475 d7579) maxlags(6) aic
> maxcombs(15000) fast

------------------------------------------------------------------------------
           w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |
         L1. |   .3346901   .0998041     3.35   0.001      .1359548    .5334255
         L2. |   .0810293   .1087734     0.74   0.459     -.1355661    .2976248
         L3. |   -.198378   .1011595    -1.96   0.053     -.3998123    .0030563
         L4. |   .4030251   .0989762     4.07   0.000      .2059382    .6001119
         L5. |  -.0693079   .0949025    -0.73   0.467     -.2582829    .1196671
         L6. |   .2017837   .0800474     2.52   0.014       .042389    .3611783
             |
        prod |   .2642678   .0587165     4.50   0.000      .1473483    .3811872
             |
          ur |
         --. |   .0038742    .008252     0.47   0.640     -.0125575     .020306
         L1. |    -.01077   .0120686    -0.89   0.375     -.0348018    .0132617
         L2. |  -.0116548   .0145987    -0.80   0.427     -.0407246    .0174151
         L3. |   .0212508   .0153697     1.38   0.171     -.0093542    .0518557
         L4. |   .0028227   .0151775     0.19   0.853     -.0273995     .033045
         L5. |  -.0304991   .0109952    -2.77   0.007     -.0523934   -.0086049
(Continued on next page)
```

Example: ARDL model with optimal lag orders

```
             |
       wedge |
         --. |  -.3059897    .051594    -5.93   0.000    -.4087265   -.2032528
         L1. |    .032749    .060641     0.54   0.591    -.0880027    .1535007
         L2. |  -.0494003   .0626044    -0.79   0.432    -.1740615    .0752609
         L3. |  -.0963634   .0611009    -1.58   0.119    -.2180308     .025304
         L4. |    .188605   .0558292     3.38   0.001      .077435    .2997751
             |
       union |
         --. |   -.955714   .8138684    -1.17   0.244    -2.576333     .664905
         L1. |  -1.467002   1.350003    -1.09   0.281    -4.155202    1.221197
         L2. |   2.527384   1.401008     1.80   0.075    -.2623798    5.317149
         L3. |    .311388   1.349422     0.23   0.818    -2.375655    2.998431
         L4. |  -2.241151   1.106961    -2.02   0.046    -4.445392   -.0369088
         L5. |   2.185799   .6535696     3.34   0.001     .8843756    3.487222
             |
       d7475 |   .0301088    .006154     4.89   0.000     .0178547     .042363
       d7579 |   .0169541   .0062481     2.71   0.008     .0045126    .0293956
       _cons |   .6604224   .1425601     4.63   0.000      .376549    .9442958
------------------------------------------------------------------------------

. matrix list e(lags)

e(lags)[1,5]
        w   prod     ur  wedge  union
r1      6      0      5      4      5
```

Example: error-correction representation

```
. ardl w prod ur wedge union if tin(1972q1,1997q4), exog(d7475 d7579) ec1 lags(6 0 5 4 5)

------------------------------------------------------------------------------
         D.w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ADJ          |
           w |
         L1. |  -.2471578   .0521006    -4.74   0.000    -.3509034   -.1434121
-------------+----------------------------------------------------------------
LR           |
        prod |
         L1. |   1.069227    .045147    23.68   0.000      .979328    1.159126
             |
          ur |
         L1. |  -.1010536   .0303893    -3.33   0.001    -.1615664   -.0405409
             |
       wedge |
         L1. |  -.9321955   .2432139    -3.83   0.000    -1.416496   -.4478946
             |
       union |
         L1. |    1.45941   .2847566     5.13   0.000      .892387    2.026433
-------------+----------------------------------------------------------------
SR           |
           w |
         LD. |  -.4181521   .0970869    -4.31   0.000    -.6114769   -.2248273
        L2D. |  -.3371228   .1076478    -3.13   0.002     -.551477   -.1227686
        L3D. |  -.5355008   .1024435    -5.23   0.000    -.7394918   -.3315098
        L4D. |  -.1324758   .0889041    -1.49   0.140    -.3095064    .0445549
        L5D. |  -.2017837   .0800474    -2.52   0.014    -.3611783    -.042389
(Continued on next page)
```

Example: error-correction representation

```
             |
        prod |
         D1. |   .2642678   .0587165     4.50   0.000     .1473483    .3811872
             |
          ur |
         D1. |   .0038742    .008252     0.47   0.640    -.0125575     .020306
         LD. |   .0180804   .0112588     1.61   0.112    -.0043387    .0404995
        L2D. |   .0064256   .0106366     0.60   0.548    -.0147545    .0276058
        L3D. |   .0276764     .01116     2.48   0.015      .005454    .0498988
        L4D. |   .0304991   .0109952     2.77   0.007     .0086049    .0523934
             |
       wedge |
         D1. |  -.3059897    .051594    -5.93   0.000    -.4087265   -.2032528
         LD. |  -.0428413   .0584202    -0.73   0.466    -.1591708    .0734882
        L2D. |  -.0922416   .0566866    -1.63   0.108     -.205119    .0206358
        L3D. |   -.188605   .0558292    -3.38   0.001    -.2997751    -.077435
             |
       union |
         D1. |   -.955714   .8138684    -1.17   0.244    -2.576333     .664905
         LD. |  -2.783421   .8141048    -3.42   0.001    -4.404511   -1.162331
        L2D. |  -.2560365   .8307344    -0.31   0.759     -1.91024    1.398167
        L3D. |   .0553516    .743211     0.07   0.941    -1.424571    1.535274
        L4D. |  -2.185799   .6535696    -3.34   0.001    -3.487222   -.8843756
             |
       d7475 |   .0301088    .006154     4.89   0.000     .0178547     .042363
       d7579 |   .0169541   .0062481     2.71   0.008     .0045126    .0293956
       _cons |   .6604224   .1425601     4.63   0.000      .376549    .9442958
------------------------------------------------------------------------------
```

Example: bounds testing

```
. estat btest

Pesaran/Shin/Smith (2001) ARDL Bounds Test
H0: no levels relationship            F =  7.367
                                      t = -4.744

Critical Values (0.1-0.01), F-statistic, Case 3

      | [I_0]   [I_1]  | [I_0]   [I_1]  | [I_0]   [I_1]  | [I_0]   [I_1]
      |   L_1     L_1  |  L_05    L_05  | L_025   L_025  |  L_01    L_01
------+----------------+----------------+----------------+---------------
  k_4 |  2.45    3.52  |  2.86    4.01  |  3.25    4.49  |  3.74    5.06
accept if F < critical value for I(0) regressors
reject if F > critical value for I(1) regressors

Critical Values (0.1-0.01), t-statistic, Case 3

      | [I_0]   [I_1]  | [I_0]   [I_1]  | [I_0]   [I_1]  | [I_0]   [I_1]
      |   L_1     L_1  |  L_05    L_05  | L_025   L_025  |  L_01    L_01
------+----------------+----------------+----------------+---------------
  k_4 | -2.57   -3.66  | -2.86   -3.99  | -3.13   -4.26  | -3.43   -4.60
accept if t > critical value for I(0) regressors
reject if t < critical value for I(1) regressors

k: # of non-deterministic regressors in long-run relationship
Critical values from Pesaran/Shin/Smith (2001)
```

- Bounds test confirms the existence of a long-run relationship.

Summary: the new `ardl` package for Stata

- The estimation of ARDL / EC models has become increasingly popular over the last decades. The associated bounds testing procedure is an attractive alternative to other cointegration tests.
- The new `ardl` command estimates an ARDL model with optimal or pre-specified lag orders.
- Two different reparameterizations of the ARDL model in EC form are available.
- The bounds procedure for testing the existence of a long-run relationship is implemented as a postestimation command. Asymptotic and finite-sample critical value bands are available.

```
net install ardl, from(http://www.kripfganz.de/stata/)
help ardl
help ardl postestimation
help ardlbounds
```

References

- Banerjee, A., J. J. Dolado, D. F. Hendry, and G. W. Smith (1986). Exploring equilibrium relationships in econometrics through static models: some Monte Carlo evidence. *Oxford Bulletin of Economics and Statistics 48(3)*: 253–277.
- Engle, R. F., and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica 55(2)*: 251–276.
- Hassler, U., and J. Wolters (2006). Autoregressive distributed lag models and cointegration. *Allgemeines Statistisches Archiv 90(1)*: 59–74.
- Narayan, P. K (2005). The saving and investment nexus for China: evidence from cointegration tests. *Applied Economics 37(17)*: 1979–1990.
- Pesaran, M. H., and Y. Shin (1998). An autoregressive distributed-lag modelling approach to cointegration analysis. In *Econometrics and Economic Theory in the 20th Century. The Ragnar Frisch Centennial Symposium*, ed. S. Strøm, chap. 11, 371–413. Cambridge: Cambridge University Press.
- Pesaran, M. H., Y. Shin, and R. Smith (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics 16(3)*: 289–326.
- Phillips, P. C. B, and B. E. Hansen (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies 57(1)*: 99–125.
- Schaffer, M. E. (2010). egranger: Stata module to perform Engle-Granger cointegration tests and 2-step ECM estimation. Statistical Software Components S457210, Boston College.