I.S.S.N: 1885-6888

ECONOMIC ANALYSIS WORKING PAPER SERIES

A Panel Data Toolbox for MATLAB

♦

Inmaculada C. Álvarez, Javier Barbero and José L. Zofío

Working Paper 05/2013

UNIVERSIDAD AUTONOMA DE MADRID

DEPARTAMENTO DE ANÁLISIS ECONÓMICO: TEORÍA ECONÓMICA E HISTORIA ECONÓMICA

A Panel Data Toolbox for MATLAB

Inmaculada C. Álvarez **Javier Barbero** **José L. Zofío**
Universidad Autónoma de Madrid

Abstract

Panel Data Toolbox is a new package for MATLAB that includes functions to estimate the main econometric methods of panel data analysis. The package includes code for the standard fixed, between and random effects estimation methods, as well as for the existing instrumental panel and new spatial panel. This paper describes the methodology and implementation of the functions and illustrates their use with well-known examples. We perform numerical checks against other popular commercial and free software in order to show the validity of the results.

Keywords: panel data, instrumental panel, spatial panel, econometrics, MATLAB.
JEL codes: C21, C23, C26.

1. Introduction

Panel data econometrics have grown in importance over the past decades due to increase availability of data related to units that are observed over several periods of time. Panel data econometric methods are available in Stata and R, but there is a lack of a full set of functions for MATLAB, by The MathWorks, Inc. (2013).

The **Panel Data Toolbox** introduces such set of functions, including estimation methods for the standard fixed, between and random effects models, as well as instrumental panel data models, including the error components by Baltagi (1981) and Baltagi and Liu (2009), and, finally, existing and new spatial panel data, Baltagi and Liu (2011). Numerical checks against Stata and R using well-known classical examples show that the estimated coefficients and t statistics are consistent with those obtained with the new MATLAB toolbox.

Spatial econometrics in MATLAB can be estimated using the LeSage and Pace (2009) **Econometrics Toolbox**, which uses maximum likelihood and bayesian methods, and Elhorst (2011) using maximum likelihood methods. In the new **Panel Data Toolbox** we use a two stage instrumental variables method to estimate spatial panels with fixed, between and random effects, as well as the error components model, following Baltagi and Liu (2011).

Panel Data Toolbox is available as free software and can be downloaded from `http://www.paneldatatoolbox.com`, with all the supplementary material (data and source code) to replicate all the results presented in this paper.

The paper is organized as follows. Section 3 presents the Panel data models with fixed, between and random effects. Instrumental panel data models are illustrated in Section 4. Spatial panels are covered in Section 5. Numerical checks against Stata and R are presented in Section 6. Finally, Section 7 concludes.

2. Data and structures

Panel data contains units (individuals, firms, countries, regions, etc.) that are observed over several periods of time. Units are usually denoted by $i = 1, 2, \ldots, n$, and time periods by $t = 1, 2, \ldots T$. In this paper we deal only with the case of balanced panel data, those in which all units are observed over the same periods of time. Then, the total number of observations in the panel is $N = nT$.

Data are managed as regular MATLAB vectors and matrices, constituting the inputs of the estimation functions. Observations are expected to be ordered first by units and then by time period. All estimation functions return a structure `estoutput` that contains properties with the estimation results as well as the input used to generate that output. Properties can be accessed directly using the dot notation and the whole structure can be used as an input to other functions that print results (e.g., `estprint`) or plot graphs (e.g., `estplot`).

Some of the properties of the `estoutput` structure are the following:[1]

- `y` and `X`: contain the dependent and the independent variables, respectively.
- `n`, `T` and `N`: number of entities, time periods, and total number of observations.
- `k` and `l`: number of explanatory variables and instruments (including the constant term).
- `coef`, `varcoef` and `stderr`: estimated coefficients, estimated covariance matrix, and estimated standard errors.
- `yhat` and `res`: fitted values and residuals.
- `statistic`, `df_statistic` and `p_statistic`: statistic of individual significance, degrees of freedom of the statistic, and the corresponding p value.

3. Panel data models

The starting formulation is the panel data model with specific individual effects:

$$y_{it} = \alpha + X_{it}\beta + \mu_i + v_{it} \qquad \forall i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{1}$$

where μ_i represents the i-th invariant time individual effect and v_{it} the disturbance, with $v_{it} \sim i.i.d(0, \theta_v^2)$, $\mathsf{E}(v_i) = 0$, $\mathsf{E}(v_i v_i^\top) = \theta_v^2 I_T$ and $\mathsf{E}(v_i v_j) = 0$ for $i \neq j$, being I_T the T x T identity matrix.

As a classic application we use Munnell (1990) and Baltagi (2008) data. Munnell (1990) suggests a Cobb-Douglas production function using data for 48 U.S. states over 17 periods (1970–1986). The dependent variable, output of the production function, is the gross state product, `log(gsp)`, and the explanatory ones are public capital, `log(pcap)`, private capital, `log(pc)`, employment, `log(emp)`, and the unemployment rate, `log(unemp)`.[2]

[1] For a full list see the help of the function typing `help estoutput` in MATLAB.

[2] Munnell (1990) data are available in MATLAB format in the supplementary file *MunnellData.mat*.

```
>> load('MunnellData.mat')

>> y = log(gsp);
>> X = [log(pcap), log(pc), log(emp), unemp];
>> T = 17;

>> dvarnames = {'lgsp'};
>> ivarnames = {'lpcap', 'lpc', 'lemp', 'unemp'};
```

We create a vector `y` containing the dependent variable and a matrix `X` with the explanatory variables. A vector of ones for the constant term should not be added to `X` because it is included internally by the estimation functions. The variables `dvarnames` and `ivarnames` are cell arrays of strings that contain the name of the variables that are subsequently used when printing the results of the estimation.

Panel data models are estimated using the `panel(y, X, T, method, options)` function, where `y` is the vector of the dependent variable, `X` is the matrix of explanatory variables, `T` is the number of time periods per entity, and `method` is a string that specifies the panel data estimation method to be used among the following:

- `po`: for a pool estimation.
- `fe`: for a fixed effects (within) estimation.
- `be`: for a between effects estimation.
- `re`: for a random effects estimation

These estimation methods are explained in the following sections. `options` is an optional parameter to specify alternative estimation choices.

3.1. Fixed effects model

Under typical specifications, individual effects are correlated with the explanatory variables: $\mathsf{COV}(X_{it}, \mu_i) \neq 0$, which motivates the use of the fixed-effects (within) estimation, so as to capture unobservable heterogeneity, Baltagi (2008).

In this context, including individual effects on the error component while performing OLS (ordinary least squares) results into a biased estimation. In order to extract these effects, the within estimator of the parameters is computed using OLS:

$$\hat{\beta}_{fe} = (\tilde{X}^\top \tilde{X})^{-1} \tilde{X}^\top \tilde{y}, \tag{2}$$

where $\tilde{y} = y - \bar{y}$ and $\tilde{X} = X - \bar{X}$ are the transformed variables in deviations from the group mean. It is called "within" estimator because it takes into account the variations in each group. This estimator is unbiased and consistent when both n and T are large. Statistical inference is generally based on the asymptotic variance covariance matrix:

$$\mathsf{VAR}(\hat{\beta}_{fe}) = S^2 (\tilde{X}^\top \tilde{X})^{-1}, \tag{3}$$

where S^2 denotes the residual variance: $S^2 = (e^\top e)/(n(T-1) - k + 1)$, with residuals $e = y - (X\hat{\beta}_{fe} + \alpha + \mu)$.

Finally, inference can be performed using the standard tests. The individual significance statistic is distributed as a t-student with $n(T-1)-k+1$ degrees of freedom under homoscedasticity, while the F statistic of joint significance is:[3]

$$F = \frac{\text{Wald}}{k-1} \sim F_{k-1,n(T-1)-k+1} \tag{4}$$

The goodness of fit is measured with the R-squared: $R^2 = 1-(e^\top e)/(\tilde{y}\tilde{y})$, and the adjusted R-squared $\bar{R}^2 = 1-(N-1)/(N-k-n)(1-R^2)$. The test for individual effects is the Chow test proposed in Baltagi (2008):

$$F = \frac{(RRSS-URSS)/(n-1)}{URSS/(n(T-1)-(k-1))} \sim F_{n-1,n(T-1)-(k-1)}, \tag{5}$$

where $RRSS$ is the restricted residual sums of squares, coming from an OLS pool estimation, and $URSS$ is the unrestricted residual sums of squares, from the fixed effects estimation.

The `panel` function implements the estimation of fixed effects panel data models in MATLAB:

```
>> regfe = panel(y, X, T, 'fe');
>> regfe.dvarnames = dvarnames;
>> regfe.ivarnames = ivarnames;
>> estprint(regfe);


Panel: Fixed effects (Within)

N observations: 816
N groups: 48
Obs per group: 17
R-squared = 0.941336
Adj R-squared = 0.941046
Joint significance: F(4, 764) = 3064.808435
                    p = 0.0000
Dept Var: lgsp
---------------------------------------------------------------
       Varname  Coefficient   Std. Error  Statistic  p-value
---------------------------------------------------------------
         lpcap     -0.02615      0.02900    -0.9017    0.368
           lpc      0.29201      0.02512    11.6246    0.000***
          lemp      0.76816      0.03009    25.5273    0.000***
         unemp     -0.00530      0.00099    -5.3582    0.000***
      CONSTANT      2.35290      0.17481    13.4595    0.000***
---------------------------------------------------------------
 Test of individual effects: F(47,764) = 75.820406
                             p-value = 0.000
```

The function `estprint` is used to display the table with the results taking the name of the variables specified in the properties `dvarnames` and `ivarnames` of the `estoutput` structure that is returned from the `panel` function.

[3] Where Wald is the standard Wald distance for joint significance tests of all estimated coefficients, excluding the constant term.

3.2. Between effects model

In the between estimation the parameters with the transformed variables:

$$\hat{\beta}_{be} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{y}, \tag{6}$$

where $\bar{y}$ and $\bar{X}$ are the means by groups, premultiplied by $\sqrt{(T)}$ to take into account that the regression is based on nT observations, since the mean of each group is repeated T times, and it should be based on the n observations, Baltagi (2008) . It is called "between" estimator because it takes into account the variation between groups, and since all observations are constant in each group. Again, this estimator is unbiased and consistent when n and T are large. Statistical inference is generally based on the asymptotic variance-covariance matrix:

$$\mathsf{VAR}(\hat{\beta}_{fe}) = S^2(\bar{X}'\bar{X})^{-1}, \tag{7}$$

where S^2 denotes the residual variance: $S^2 = (e^\top e)/(n-k)$, with residuals $e = y - \bar{X}\hat{\beta}_{fe}$. The statistic of individual significance is distributed as a t student with $n-k$ degrees of freedom. The Wald distance is computed as usual and the F statistic of joint significance is:

$$F = \frac{\text{Wald}}{k-1} \sim F_{k-1,n-k}. \tag{8}$$

The goodness of fit is measured with the R^2, which is computed as the square of the correlation coefficient of $\bar{y}$ and $\hat{y}$.

The `panel` function implements the estimation of between effects panel data in MATLAB:

```
>> regbe = panel(y, X, T, 'be');
>> regbe.dvarnames = dvarnames;
>> regbe.ivarnames = ivarnames;
>> estprint(regbe);
```

```
Panel: Between effects

N observations: 816
N groups: 48
Obs per group: 17
R-squared = 0.993909
Joint significance: F(4, 43) = 1754.114154
                    p = 0.0000
Dept Var: lgsp
```

Varname	Coefficient	Std. Error	Statistic	p-value
lpcap	0.17937	0.07197	2.4922	0.017**
lpc	0.30195	0.04182	7.2201	0.000***
lemp	0.57613	0.05637	10.2196	0.000***
unemp	−0.00389	0.00991	−0.3926	0.697
CONSTANT	1.58944	0.23298	6.8222	0.000***

3.3. Random effects model

In the panel data model (1) the loss of degrees of freedom can be avoided if the individual effects can be assumed random, where the error component $u_{it} = \mu_i + v_{it}$ includes the i-th invariant time individual effects μ_i and the disturbance v_{it}.

$$y_{it} = \alpha + X_{it}\beta + u_{it} \qquad \forall i = 1, \dots, n, \quad t = 1, \dots, T \tag{9}$$

The individual effect μ_i is assumed independent of the disturbance v_{it}. In addition, individual effects and disturbances are independent of the explanatory variables, i.e., $\mathsf{COV}(X_{it}, \mu_i) \neq 0$ and $\mathsf{COV}(X_{it}, v_{it}) \neq 0$ for all i and t. For this reason, the random effects model is an appropriate specification in the analysis of n individuals randomly drawn from a large population. In this context, n is usually large and a fixed effects model would lead to a loss of degrees of freedom.

From the composed error component,

$$\mathsf{E}(\mu_i) = \mathsf{E}(v_{it}) = \mathsf{E}(\mu_i v_{it}) = 0 \tag{10}$$

$$\mathsf{E}(\mu_i \mu_j) = \begin{cases} \sigma_\mu^2 & i \neq j \\ 0 & i = j \end{cases} \qquad \mathsf{E}(v_i v_j) = \begin{cases} \sigma_v^2 & i \neq j \\ 0 & i = j \end{cases} \tag{11}$$

This results in a block-diagonal covariance matrix with serial correlation over time only between disturbances of the same individual and zero otherwise:

$$\mathsf{COV}(u_{it}, u_{js}) = \begin{cases} \sigma_\mu^2 + \sigma_v^2 & i = j, t = s \\ \sigma_\mu^2 & i = j, t \neq s \end{cases} \tag{12}$$

This implies the following correlation coefficient between disturbances:

$$\rho = \mathsf{CORR}(u_{it}, u_{js}) = \begin{cases} 1 & i = j, t = s \\ \sigma_\mu^2/(\sigma_\mu^2 + \sigma_v^2) & i = j, t \neq s \end{cases} \tag{13}$$

Therefore, the covariance matrix can be computed as follows:

$$\Omega = \mathsf{E}(uu^\top) = \sigma_\mu^2(I_n \otimes J_T) + \sigma_v^2(I_n \otimes I_T), \tag{14}$$

where J_T is a matrix of ones of size T and the homoscedastic variance is $\mathsf{VAR}(u_{it}) = \sigma_\mu^2 + \sigma_v^2$ for all i and t. In this case, the GLS (generalized least squares) method yields an efficient estimator of the parameters. Following the general expression (White, 1980),

$$\hat{\beta}_{re} = (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} y, \tag{15}$$

with $\Omega^{-1} = 1/\sigma_v^2 \left[I_T - \sigma_\mu^2/(\sigma_v^2 + T\sigma_\mu^2)\right]$. In order to obtain the GLS estimator of the regression coefficients, it is necessary to estimate Ω^{-1} which is a matrix of dimension nT x nT. The GLS estimation of the random effects model is based on the transformation proposed by Baltagi (2008):

$$\hat{\beta}_{re} = (\tilde{X}^\top \tilde{X})^{-1} \tilde{X}^\top y, \tag{16}$$

where $\tilde{y} = y - \theta\bar{y}$ and $\tilde{X} = X - \theta\bar{X}$ are the transformed variables in quasideviations from the group mean. The factor *theta* corresponds to Greene (2012):

$$\theta = 1 - \sqrt{\frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\mu^2}}. \tag{17}$$

Focusing on a different derivation based on the spectral decomposition of Ω one obtains, Baltagi (2008):

$$\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2 \tag{18}$$

The random effects estimator (16) is a weighted average of the within and between estimators, with the ratio $\sigma_v^2/(\sigma_v^2 + T\sigma_\mu^2)$ being the weight assigned to the between groups variation. Therefore, under the assumption of fixed effects this latter variation is omitted, with the ratio equal to zero and θ equal to one (opposite to the OLS case). As a result, the treatment of individual effects as random provides an intermediate solution between complete variation and time invariant fixed effects.

Swamy and Arora (1972) suggest using the within regression residuals to compute σ_v^2 and the residuals from the between regression to compute σ_1^2. From these estimates σ_μ^2 can be calculated as: [4].

$$\sigma_\mu^2 = \frac{\sigma_1^2 - \sigma_v^2}{T} \tag{19}$$

Statistical inference is generally based on the asymptotic variance-covariance matrix:

$$\mathsf{VAR}(\hat{\beta}_{re}) = S^2(\tilde{X}^\top\tilde{X})^{-1}, \tag{20}$$

where, once again, S^2 denotes the residual variance: $S^2 = (e^\top e)/(N-k)$, with residuals $e = y - \tilde{X}\hat{\beta}_{re}$.

Finally, the statistic of individual significance is computed as usual and it is normally distributed. Also, Wald distance for joint significance is computed as before, and the statistic of joint significance is:

$$\chi^2 = Wald \sim \chi^2_{k-1}. \tag{21}$$

The goodness of fit is measured with the R^2, which is computed as the square of the correlation coefficient of $\hat{y}$ and $\tilde{y}$.

The `panel` function implements the estimation of random effects panel data in MATLAB:

```
>> regre = panel(y, X, T, 're');
>> regre.dvarnames = dvarnames;
>> regre.ivarnames = ivarnames;
>> estprint(regre);
```

[4] If the estimated σ_μ^2 is negative, which occurs when the true value is closed to zero (Baltagi 2008, p. 20), it may be replaced by zero as suggested by Maddala and Mount (1973)

```
Panel: Random effects GLS (Swamy and Arora)

N observations: 816
N groups: 48
Obs per group: 17
R-squared = 0.959332
Joint significance: Chi2(4) = 19131.085009
                    p-value = 0.0000
Dept Var: lgsp
```

Varname	Coefficient	Std. Error	Statistic	p-value
lpcap	0.00444	0.02342	0.1895	0.850
lpc	0.31055	0.01980	15.6805	0.000***
lemp	0.72967	0.02492	29.2803	0.000***
unemp	-0.00617	0.00091	-6.8033	0.000***
CONSTANT	2.13541	0.13346	16.0002	0.000***

```
 sigma_mu = 0.082691
  sigma_v = 0.038137     sigma_1 = 0.343068
   rho_mu = 0.824601       Theta = 0.888835
```

Here `rho_mu` is the fraction of variance due to the individual effects and it is computed as $\rho_\mu = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_v^2)$.

3.4. Hausman test of specification

In order to determine the correct specification of the model, fixed versus random effects, it is necessary to check the correlation between the individual effects and the regressors. When the individuals effects and the explanatory variables are correlated: $E(\mu_i X_{it}) \neq 0$, the fixed effects model provides an unbiased estimator, otherwise a feasible GLS is an efficient estimator in a random effects model.

Hausman (1978) suggests comparing the GLS estimator of the random effects model $\hat{\beta}_{re}$ and the within estimator in the fixed effects model $\hat{\beta}_{fe}$, both of which are consistent under the null hypothesis $H_0 : E(\mu_i X_{it}) = 0$. Under H_0 the GLS estimator is BLUE, consistent and asymptotically efficient, while the within estimator is consistent whether H_0 is true or not. Furthermore, the GLS estimator is inconsistent if H_0 is false. Therefore, the statistic would be based on the difference between both estimators: $\hat{\beta}_{fe} - \hat{\beta}_{re}$.

Hence, the Hausman test statistic is given by (Baltagi (2008)):

$$H = (\hat{\beta}_{fe} - \hat{\beta}_{re})^\top \mathsf{VAR}(\hat{\beta}_{fe} - \hat{\beta}_{re})^{-1}(\hat{\beta}_{fe} - \hat{\beta}_{re}) \sim \chi^2_{k-1}, \tag{22}$$

where $\mathsf{VAR}(\hat{\beta}_{fe} - \hat{\beta}_{re})^{-1} = \mathsf{VAR}(\hat{\beta}_{fe}) - \mathsf{VAR}(\hat{\beta}_{re})$.

For n fixed and T large, both estimators tend to similar values, with their difference converging to zero, and Hausman's test is unnecessary. However, in applications where n is relatively large with respect to T, it can be used to choose between estimators.

The `[H, p] = hausman(estA, estB)` function implements the Hausman test in MATLAB, where the input arguments `estA` and `estB` are `estoutput` structures of the previous estimations. The function returns the value of the test, `H`, and its corresponding p value, `p`. To display the results in a table, the `hausmanprint(estA, estB)` must be used:

```
>> hausmanprint(regfe, regre);

Hausman's test of specification
```

Varname	EstimationA Panel−fe	EstimationB Panel−re	Coef Difference	Std. Error Difference
lpcap	−0.026150	0.004439	−0.030588	0.01711
lpc	0.292007	0.310548	−0.018542	0.01545
lemp	0.768159	0.729671	0.038489	0.01687
unemp	−0.005298	−0.006172	0.000875	0.00039

```
estA is consistent under H0 and H1 (estA = Panel−fe)
estB is consistent under H0        (estB = Panel−re)
H0: coef(estA) − coef(estB) = 0
H1: coef(estA) − coef(estB) != 0
      H = 9.5254156
p−value = 0.04923
H is distributed as Chi2(4)
```

3.5. Heteroscedasticity in panel data models

The fixed effects model can be estimated using the within estimator and a robust covariance matrix when the disturbances are affected by heteroscedasticity. Hansen (2007) proposed a robust estimation of the parameters' covariance matrix using the White sandwich estimator, White (1980):

$$\mathsf{VAR}(\hat{\beta}_{fe}) = \frac{n}{n-1}\frac{N-1}{N-k}(\tilde{X}^\top\tilde{X})^{-1}\left[\sum_{i=1}^{n}\tilde{X}_i^\top e_i e_i^\top \tilde{X}_i\right](\tilde{X}^\top\tilde{X})^{-1} \tag{23}$$

From $\mathsf{VAR}(\hat{\beta}_{fe})$ the correct variance of the constant term must be computed as:

$$\mathsf{VAR}(\hat{\alpha}_{fe}) = (\frac{1}{n}\sum_{i=1}^{n}(\frac{1}{T}\sum_{t=1}^{T}X_{it}))\mathsf{VAR}(\hat{\beta}_{fe})(\frac{1}{n}\sum_{i=1}^{n}(\frac{1}{T}\sum_{t=1}^{T}X_{it}))^\top \tag{24}$$

For the random effects model the robust estimation of the parameters' covariance matrix is computed using an estimator equivalent to that proposed by White (1980), (23), but with the suitable transformation of the variables.

The `panel` function, with the `options` argument set to `robust`, implements the estimation of fixed effects robust panel data models in MATLAB:

```
>> regfer = panel(y, X, T, 'fe', 'robust');
>> regfer.dvarnames = dvarnames;
>> regfer.ivarnames = ivarnames;
>> estprint(regfer);
```

```
Panel: Fixed effects (Within)

N observations: 816
N groups: 48
Obs per group: 17
R-squared = 0.941336
Adj R-squared = 0.941046
Joint significance: F(4, 47) = 395.610133
                    p = 0.0000
Dept Var: lgsp
Robust standard errors adjusted for 48 clusters
```

Varname	Coefficient	Std. Error	Statistic	p-value
lpcap	-0.02615	0.06111	-0.4279	0.671
lpc	0.29201	0.06255	4.6684	0.000***
lemp	0.76816	0.08273	9.2848	0.000***
unemp	-0.00530	0.00253	-2.0952	0.042**
CONSTANT	2.35290	0.31459	7.4792	0.000***

In the random effects model the **robust** option provides the robust covariance matrix estimation:

```
>> regrer = panel(y, X, T, 're', 'robust');
>> regrer.dvarnames = dvarnames;
>> regrer.ivarnames = ivarnames;
>> estprint(regrer);
```

```
Panel: Random effects GLS (Swamy and Arora)

N observations: 816
N groups: 48
Obs per group: 17
R-squared = 0.959332
Joint significance: Chi2(4) = 4408.644223
                    p-value = 0.0000
Dept Var: lgsp
Robust standard errors adjusted for 48 clusters
```

Varname	Coefficient	Std. Error	Statistic	p-value
lpcap	0.00444	0.05531	0.0802	0.936
lpc	0.31055	0.04416	7.0320	0.000***
lemp	0.72967	0.07088	10.2941	0.000***
unemp	-0.00617	0.00236	-2.6120	0.009***
CONSTANT	2.13541	0.24179	8.8318	0.000***

```
 sigma_mu = 0.082691
  sigma_v = 0.038137     sigma_1 = 0.343068
   rho_mu = 0.824601       Theta = 0.888835
```

4. Instrumental panel data models

The assumption of exogeneity of the independent variables, X, when they are uncorrelated with the disturbance, $E(X_{it}, v_{it}) = 0$, implies that OLS remains valid. However, there are many applications in which this assumption is untenable. In this case, when the regressors are endogenous, the OLS estimator loses consistency and unbiasedness. Consequently, we can apply an instrumental variables (IV) two stage estimation to the fixed effects, random effects and between models, Greene (2012).

We assume that there is a set of variables that are exogenous, uncorrelated with the disturbance, and relevant, i.e., correlated with the endogenous independent variables. This set is represented by the H matrix.

For an application of instrumental panel data, we follow Baltagi and Levin (1992) and Baltagi, Griffin, and Xiong (2000) who estimate the demand for cigarettes using data from 46 U.S. states over the period 1963–1992.[5] We estimate the consumption, `c`, measured as per capita sales, which depends on the price per pack, `price`, per capita disposable income, `ndi`, and the minimum price in neighbor states, `pimin`. The instruments normally used are the lags of the disposable income, `ndi_1`, and the lag of the minimum price `pimin_1`.[6]

```
>> load('CigarData.mat')

>> y = log(c);
>> X = [log(price), log(ndi), log(pimin)];
>> H = [log(ndi_1), log(pimin_1), log(ndi), log(pimin)];
>> T = 29;

>> dvarnames = {'lc'};
>> ivarnames = {'lprice', 'lndi', 'lpimin'};
```

Instrumental panel data models are estimated using the `ivpanel(y, X, H, T, method)` function, where `y` is the vector of the dependent variable, `X` is the matrix of explanatory variables, `H` is the matrix of instruments, `T` is the number of time periods per unit, and `method` is a string that specifies the choice of panel data estimation method, among the following:

- `po`: for a pool estimation.
- `fe`: for a fixed effects (within) estimation.
- `be`: for a between effects estimation.
- `re`: for a random effects estimation
- `ec`: for a error-components estimation, Baltagi and Liu (2009).

[5] Data is available in MATLAB format in the supplementary file *CigarData.mat*.

[6] The equation we estimate differs from the original one, which corresponds to a dynamic panel data model.

4.1. Two stage least squares (2SLS)

The first stage of the 2SLS estimation consists of estimating the independent variables, $\hat{X}$, by an OLS estimate of X over the exgonenous variables and instruments, H:[7]

$$\hat{X} = \tilde{H}(\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top \tilde{X} \tag{25}$$

The second stage consists in estimating the coefficients, $\hat{\beta}$, using the predicted $\hat{X}$:

$$\hat{\beta}_{2SLS} = (\hat{X}^\top \hat{X})^1 \hat{X}^\top \tilde{y} \tag{26}$$

In each case, $\tilde{y}$, $\tilde{X}$ and $\tilde{H}$ represent the different transformations applied to the variables to obtain the within, between and GLS estimator as explained in Section 3. Regarding statistical inference, the statistic of individual significance is normally distributed, while the statistic of joint significance is distributed as a χ^2 with $k-1$ degrees of freedom. The test for individual effects is that proposed in Baltagi (2008).

The **ivpanel** function implements the estimation of fixed, between and random effects instrumental panel data models in MATLAB:

```
>> regivfe = ivpanel(y, X, H, T, 'fe');
>> regivfe.dvarnames = dvarnames;
>> regivfe.ivarnames = ivarnames;
>> estprint(regivfe)
```

```
IV Panel: Fixed effects (Within)

N observations: 1334
N groups: 46
Obs per group: 29
R-squared = 0.640642
Joint significance: Chi2(3) = 1792.756633
                    p-value = 0.0000
Dept Var: lc
```

Varname	Coefficient	Std. Error	Statistic	p-value
lprice	-1.01636	0.24920	-4.0785	0.000***
lndi	0.53785	0.02303	23.3507	0.000***
lpimin	0.31237	0.22839	1.3677	0.171
CONSTANT	2.99141	0.08111	36.8827	0.000***

```
>> regivbe = ivpanel(y, X, H, T, 'be');
>> regivbe.dvarnames = dvarnames;
>> regivbe.ivarnames = ivarnames;
>> estprint(regivbe)
```

[7] Note that the matrix H must include the instruments as well as the exogenous variables that are also included in X, which are instruments of themselves.

```
IV Panel: Between effects

N observations: 1334
N groups: 46
Obs per group: 29
R-squared = 0.311151
Joint significance: Chi2(3) = 6.660389
                    p-value = 0.0835
Dept Var: lc
```

Varname	Coefficient	Std. Error	Statistic	p-value
lprice	-3.27523	2.61392	-1.2530	0.210
lndi	0.83220	0.40039	2.0785	0.038**
lpimin	1.18107	1.32375	0.8922	0.372
CONSTANT	6.17390	3.29673	1.8727	0.061*

```
>> regivre = ivpanel(y, X, H, T, 're');
>> regivre.dvarnames = dvarnames;
>> regivre.ivarnames = ivarnames;
>> estprint(regivre)
```

```
IV Panel: Random effects GLS (Swamy and Arora)

N observations: 1334
N groups: 46
Obs per group: 29
R-squared = 0.638272
Joint significance: Chi2(3) = 1820.426405
                    p-value = 0.0000
Dept Var: lc
```

Varname	Coefficient	Std. Error	Statistic	p-value
lprice	-1.00711	0.24735	-4.0715	0.000***
lndi	0.53747	0.02303	23.3398	0.000***
lpimin	0.30357	0.22643	1.3407	0.180
CONSTANT	2.99212	0.08567	34.9268	0.000***

```
 sigma_mu = 0.190101
  sigma_v = 0.077566     sigma_1 = 1.026661
   rho_mu = 0.857278       Theta = 0.924449
```

4.2. Error components two stage least squares (EC2SLS)

Baltagi (1981) and Baltagi and Liu (2009) propose a generalized two stage least squares (G2SLS) estimation using the following matrix of instruments:

$$A = \left[\tilde{H}, \bar{H}\right], \tag{27}$$

where $\tilde{H}$ contains the transformed instruments in deviations from the group mean, and $\bar{H}$ the group means. The 2SLS estimation is then performed using this matrix of instruments.[8]

The error components two stage least squares (EC2SLS) estimator is consistent and presents the same limiting distribution than the G2SLS estimator. Although it is worth noting that for small samples the former shows gains in efficiency, Baltagi and Liu (2009).

The `ivpanel` function provides an estimation of the error components two stage least squares (EC2SLS) model in MATLAB by specifying the `ec` method:

```
>> regivec = ivpanel(y, X, H, T, 'ec');
>> regivec.dvarnames = dvarnames;
>> regivec.ivarnames = ivarnames;
>> estprint(regivec)
```

```
IV Panel: Error components (EC2SLS)

N observations: 1334
N groups: 46
Obs per group: 29
R-squared = 0.638782
Joint significance: Chi2(3) = 1825.252894
                    p-value = 0.0000
Dept Var: lc
```

Varname	Coefficient	Std. Error	Statistic	p-value
lprice	-0.99268	0.23587	-4.2086	0.000***
lndi	0.53641	0.02236	23.9939	0.000***
lpimin	0.29039	0.21597	1.3446	0.179
CONSTANT	2.99512	0.08420	35.5724	0.000***

```
 sigma_mu = 0.190101
  sigma_v = 0.077566     sigma_1 = 1.026661
   rho_mu = 0.857278       Theta = 0.924449
```

5. Spatial panel data models

In recent years the econometrics literature has grown with topics related to the analysis of spatial relations using panel data models. The main reason is the availability of more complete data sets in which units characterized by spatial features are followed over time. In general, a spatial panel data set contains more information and less multicollinearity among the variables

[8] The instruments A are used in the 2SLS procedure, but only H is used when estimating σ_v^2 and σ_1^2.

than a cross-section spatial counterpart (see Anselin (1988, 2010) for an introduction to this literature). Additionally, the use of panel data increases the efficiency due to larger degrees of freedom and allows the inclusion of unobservable heterogeneities Baltagi (2008).

In the context of cross-sectional models, Kelejian and Prucha (1998) introduced a generalized spatial two-stage least squares estimator, Kelejian and Prucha (1999)[9] proposed a generalized moments (GM) estimation method feasible even when n is large, while Anselin (1988) provided the ML (Maximum likelihood) estimator. Kapoor, Kelejian, and Prucha (2007) generalized the GM procedure from cross-section to panel data and derived its properties when T is fixed and n tends to infinite. Most recently, Elhorst (2003, 2010) and Lee and Yu (2010) presented the ML estimators of the spatial lag model as well as the error model extended to include fixed and random effects, solving the computational problems when the number of cross sectional units n is large. In line with Anselin (1988) and Kapoor *et al.* (2007), Baltagi, Egger, and Pfaffermayr (2006) suggest a generalized spatial panel model allowing for spatial correlation in the individual and the remainder error components. They derive the ML estimator for this more general spatial panel model with random effects.

In order to compute different estimators in spatial panel models, we consider the Cliff-Ord autoregressive spatial panel model:

$$y_{it} = \lambda W y_{it} + \beta_X X + \beta_L W L + u_i + v_{it}, \tag{28}$$

where the matrix L contains the spatial lagged independent variables, which usually are also included in X.

The application is based on Munnell (1990) and Baltagi (2008) data of U.S. states production as in Section 3.[10]

```
>> load('MunnellData.mat')

>> y = log(gsp);
>> X = [log(pcap), log(pc), log(emp), unemp];
>> L = [ ];
>> T = 17;

>> dvarnames = {'lgsp'};
>> ivarnames = {'lpcap', 'lpc', 'lemp', 'unemp'};

>> load('MunnellW.mat');

>> W = kron(W, eye(T));
```

We use the kronecker product to replicate the W adjacency matrix of the 48 U.S. states over all time periods:

$$W_{big} = W \otimes I_T \tag{29}$$

[9] Kelejian and Prucha (2004) extend the model to a system of equation spatially interrelated, while Kelejian and Prucha (2007, 2010) introduce a method robust to heteroscedasticity and autocorrelation in disturbances in a spatial autoregressive model.

[10] Munnel (1980) data is available in MATLAB format in the supplementary file `MunnellData.mat`, and the W matrix in the file `MunnellW.mat`.

Spatial panel data models are estimated using the `spanel(y, X, L, W, T, method)` function, where `y` is the vector of the dependent variable, `X` is the matrix of explanatory variables, `L` is the matrix of spatial lagged independent variables, `T` is the number of time periods per unit, and `method` is a string that specifies the panel data estimation method to use, among the following:

- `po`: for a spatial pool estimation.
- `fe`: for a spatial fixed effects (within) estimation.
- `be`: for a spatial between effects estimation.
- `re`: for a spatial random effects estimation
- `ec`: for a spatial error components estimation, Baltagi and Liu (2011).
- `sec-b`: for a spatial error components best estimation, Baltagi and Liu (2011).

5.1. Generalized two stage least squares (GS2SLS)

The spatial panels are computed as an instrumental variable estimation, extending the generalized spatial two stage least squares estimator (GS2SLS) provided by Kelejian and Prucha (1998) with fixed, between and random effects.

For simplicity, we rewrite the model more compactly as follows:

$$y_{it} = \delta Z_{it} + u_{it}, \tag{30}$$

where $Z_{it} = (Wy_{iy}, X_{it}, WL_{it})$ and $\delta = (\lambda, \beta_X, \beta_L)$. Following Kelejian and Prucha (1998) we build the matrix of instruments as:

$$H = \left[X, WX, W^2X\right] \tag{31}$$

We compute the first stage of the GS2SLS method estimating the fitted values for the independent variables $\hat{Z}$ performing OLS of Z on the instruments H:

$$\hat{Z} = \tilde{H}(\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top \tilde{Z}. \tag{32}$$

In the second stage we compute the coefficients, $\hat{\delta}$, using the predicted $\hat{Z}$:

$$\hat{\delta} = (\hat{Z}^\top \hat{Z})^{-1}\hat{Z}\tilde{y} \tag{33}$$

In each case, $\tilde{y}$, $\tilde{X}$ and $\tilde{Z}$ represent the different transformations applied to their corresponding set of variables to obtain the alternative estimations: fixed effects spatial two stage least squares (FE-S2SLS), between effects spatial two stage least squares (BE-2SLS), and random effects spatial two stage least squares (RE-S2SLS).

The fitted values are computed as in Elhorst (2003, 2010):

$$\hat{y} = (I_N - \lambda W)^{-1}(X\beta_X + WL\beta_L) \tag{34}$$

The **spanel** function implements the estimation of the fixed, between, random and error components spatial panel data models in MATLAB:

```
>> regsfe = spanel(y, X, L, W, T, 'fe');
>> regsfe.dvarnames = dvarnames;
>> regsfe.ivarnames = ivarnames;
>> estprint(regsfe);


Spatial Panel: Fixed effects (FE-2SLS)

N observations: 816
N groups: 48
Obs per group: 17
Joint significance: Chi2(5) = 12845.284388
                    p-value = 0.0000
Dept Var: lgsp
```

Varname	Coefficient	Std. Error	Statistic	p-value
lpcap	−0.04041	0.02846	−1.4197	0.156
lpc	0.21904	0.02679	8.1770	0.000***
lemp	0.66833	0.03285	20.3447	0.000***
unemp	−0.00473	0.00097	−4.8683	0.000***
W*lgsp	0.19166	0.02794	6.8597	0.000***
CONSTANT	1.93735	0.18150	10.6741	0.000***

```
>> regsbe = spanel(y, X, L, W, T, 'be');
>> regsbe.dvarnames = dvarnames;
>> regsbe.ivarnames = ivarnames;
>> estprint(regsbe);


Spatial Panel: Between effects (BE-2SLS)

N observations: 816
N groups: 48
Obs per group: 17
Joint significance: Chi2(5) = 6901.939058
                    p-value = 0.0000
Dept Var: lgsp
```

Varname	Coefficient	Std. Error	Statistic	p-value
lpcap	0.17131	0.07487	2.2882	0.022**
lpc	0.30163	0.04217	7.1520	0.000***
lemp	0.58559	0.06082	9.6283	0.000***
unemp	−0.00242	0.01054	−0.2297	0.818
W*lgsp	−0.01082	0.02474	−0.4373	0.662
CONSTANT	1.70896	0.36036	4.7423	0.000***

```
>> regsre = spanel(y, X, L, W, T, 're');
>> regsre.dvarnames = dvarnames;
>> regsre.ivarnames = ivarnames;
>> estprint(regsre);


Spatial Panel: Random effects (RE-2SLS)

N observations: 816
N groups: 48
Obs per group: 17
Joint significance: Chi2(5) = 18847.914957
                    p-value = 0.0000
Dept Var: lgsp
--------------------------------------------------------------
       Varname  Coefficient   Std. Error  Statistic  p-value
--------------------------------------------------------------
         lpcap      0.02286     0.02492     0.9173    0.359
           lpc      0.29376     0.02104    13.9596    0.000***
          lemp      0.70864     0.02679    26.4514    0.000***
         unemp     -0.00648     0.00092    -7.0525    0.000***
        W*lgsp      0.03547     0.01481     2.3946    0.017**
      CONSTANT      1.90996     0.16628    11.4865    0.000***
--------------------------------------------------------------
 sigma_mu = 0.083405
  sigma_v = 0.037325     sigma_1 = 0.345907
   rho_mu = 0.833143       Theta = 0.892094


>> regsec = spanel(y, X, L, W, T, 'ec');
>> regsec.dvarnames = dvarnames;
>> regsec.ivarnames = ivarnames;
>> estprint(regsec);


Spatial Panel: Error Components (SEC-2SLS)
N observations: 816
N groups: 48
Obs per group: 17
Joint significance: Chi2(5) = 18842.897203
                    p-value = 0.0000
Dept Var: lgsp
--------------------------------------------------------------
       Varname  Coefficient   Std. Error  Statistic  p-value
--------------------------------------------------------------
         lpcap      0.02454     0.02491     0.9850    0.325
           lpc      0.29239     0.02104    13.8978    0.000***
          lemp      0.70670     0.02678    26.3879    0.000***
         unemp     -0.00651     0.00092    -7.0840    0.000***
        W*lgsp      0.03850     0.01476     2.6089    0.009***
      CONSTANT      1.89001     0.16608    11.3803    0.000***
--------------------------------------------------------------
 sigma_mu = 0.083405
  sigma_v = 0.037325     sigma_1 = 0.345907
   rho_mu = 0.833143       Theta = 0.892094
```

5.2. Spatial error components best two stage least squares (SEC-B2SLS)

Baltagi and Liu (2011) extend the error component two-stage least square estimator proposed by Baltagi (1981), following the method introduced by Kelejian and Prucha (1998) and using Lee (2003) optimal instrument for this spatial autoregressive panel model. They obtain the spatial error components best two stage least squares estimator (SEC-B2SLS), in which we base our estimation.

Accordingly, we consider the following matrix of instruments:

$$B = \left[\tilde{H}_b^{\star}, \bar{H}_b^{\star}\right], \tag{35}$$

where $\tilde{H}_b^{\star} = \left[\tilde{X}, WA^{-1}\tilde{X}\beta\right]$ and $\bar{H}_b^{\star} = \left[\bar{X}, WA^{-1}\bar{X}\beta\right]$ are the instruments with the transformations used in the fixed and between models, respectively, and $A = (I_N - \lambda W)$. λ and β are consistent estimators and can be those obtained from a pool spatial regression. Then, GSL estimation is performed using the matrix of instruments B.

```
>> regsecb = spanel(y, X, L, W, T, 'sec-b');
>> regsecb.dvarnames = dvarnames;
>> regsecb.ivarnames = ivarnames;
>> estprint(regsecb);
```

```
Spatial Panel: Error Components Best (SEC-B2SLS)

N observations: 816
N groups: 48
Obs per group: 17
R-squared = 0.958805
Joint significance: Chi2(5) = 18852.569351
                    p-value = 0.0000
Dept Var: lgsp
```

Varname	Coefficient	Std. Error	Statistic	p-value
lpcap	0.01838	0.02493	0.7371	0.461
lpc	0.29742	0.02105	14.1284	0.000***
lemp	0.71380	0.02680	26.6310	0.000***
unemp	-0.00640	0.00092	-6.9677	0.000***
W*lgsp	0.02739	0.01489	1.8396	0.066*
CONSTANT	1.96319	0.16656	11.7869	0.000***

```
 sigma_mu = 0.083405
  sigma_v = 0.037325     sigma_1 = 0.345907
   rho_mu = 0.833143       Theta = 0.892094
```

6. Numerical checks

Numerical checks against other commercial and free software are performed by comparing the standard panel data results obtained in Section 3 from this **Panel Data Toolbox** in MATLAB and the results reported by Stata, `xtreg` function, and the R package **plm** by Croissant and Millo (2008), `plm` function.

Results for the fixed, between and random estimators using the Munnell (1990) data are reported in Table 1. The decimal places are those corresponding to the default output of all three softwares.[11] Results show that there are not differences in the estimated coefficients and t-statistics between the three programs.[12]

		Coefficient			t statistic		
		MATLAB	Stata	R	MATLAB	Stata	R
Fixed	lpcap	-0.02615	-.0261493	-0.02614965	-0.9017	-0.90	-0.9017
	lpc	0.29201	.2920067	0.29200693	11.6246	11.62	11.6246
	lemp	0.76816	.7681595	0.76815947	25.5273	25.53	25.5273
	unemp	-0.00530	-.0052977	-0.00529774	-5.3582	-5.36	-5.3582
	CONST	2.35290	2.352898	N.A.	13.4595	13.46	N.A.
Between	lpcap	0.17937	.1793651	0.1793651	2.4922	2.49	2.4922
	lpc	0.30195	.3019542	0.3019542	7.2201	7.22	7.2201
	lemp	0.57613	.5761274	0.5761274	10.2196	10.22	10.2196
	unemp	-0.00389	-.0038903	-0.0038903	-0.3926	-0.39	-0.3926
	CONST	1.58944	1.589444	1.5894444	6.8222	6.82	6.8222
Random	lpcap	0.00444	.0044388	0.00443859	0.1895	0.19	0.1895
	lpc	0.31055	.3105483	0.31054843	15.6805	15.68	15.6805
	lemp	0.72967	.7296705	0.72967053	29.2803	29.28	29.2803
	unemp	-0.00617	-.0061725	-0.00617247	-6.8033	-6.80	-6.8033
	CONST	2.13541	2.135411	2.13541100	16.0002	16.00	16.0002

Table 1: Comparison of estimated coefficients and t statistics for panel data against Stata and R.

Checks for the instrumental variables panel data models with fixed, between, random, and error components for Stata, using the `xtivreg` function, and R package **plm** function by Croissant and Millo (2008), are reported in Table 2, using the cigarette data, Baltagi (2008).[13] Again, results are the the same for all three programs.

Spatial panel estimations are checked against the R package **splm** by Millo and Piras (2012)[14], using the `spgm` function, which performs a GM implementation.[15]. Results in Table 3 reveal slight differences in the estimated coefficients and t statistics, but these differences do not change the overall features of the estimation results.

[11] All decimals can be obtained for the **Panel Data Toolbox** accessing directly the properties `coef` or `statistic` of the `estoutput` structure.

[12] The code is available in the supplementary files `NC_panel_Stata.do` and `NC_panel_R.R`.

[13] The code is available in the supplementary files `NC_ivpanel_Stata.do` and `NC_ivpanel_R.R`.

[14] The R package **sphet** by Piras (2010) can estimate spatial models with heteroskedastic innovations.

[15] The code is available in the supplementary file `NC_spanel_R.R`.

		Coefficient			*t* statistic		
		MATLAB	Stata	R	MATLAB	Stata	R
Fixed	lprice	-1.01636	-1.016359	-1.016355	-4.0785	-4.08	-4.0785
	lndi	0.53785	.5378483	0.537848	23.3507	23.35	23.3507
	lpimin	0.31237	.3123759	0.312372	1.3677	1.37	1.3677
	CONST	2.99141	2.99141	N.A.	36.8827	36.88	N.A.
Between	lprice	-3.27523	-3.275225	-3.27523	-1.2530	-1.25	0.21714
	lndi	0.83220	.8322024	0.83220	2.0785	2.08	0.04381
	lpimin	1.18107	1.181067	1.18107	0.8922	0.89	0.37736
	CONST	6.17390	6.173898	6.17390	1.8727	1.87	1.8727
Random	lprice	-1.00711	-1.007117	-1.007113	-4.0715	-4.07	-4.0715
	lndi	0.53747	.5374736	0.537473	23.3398	23.34	23.3398
	lpimin	0.30357	.3035705	0.303567	1.3407	1.34	1.3407
	CONST	2.99212	2.992121	2.992121	34.9268	34.93	34.9268
Error components	lprice	-0.99268	-.9926806	-0.992679	-4.2086	-4.21	-4.2086
	lndi	0.53641	.5364105	0.536410	23.9939	23.99	23.9939
	lpimin	0.29039	.2903891	0.290388	1.3446	1.34	1.3446
	CONST	2.99512	2.995124	2.995124	35.5724	35.57	35.5724

Table 2: Comparison of estimated coefficients and t statistics for instrumental panel data against Stata and R.

		Coefficient		*t* statistic	
		MATLAB	R	MATLAB	R
Fixed	lpcap	-0.04041	-0.03994235	-1.4197	-1.4952
	lpc	0.21904	0.22141443	8.1770	8.7866
	lemp	0.66833	0.67158117	20.3447	21.7199
	unemp	-0.00473	-0.00474680	-4.8683	-5.2069
	W*lgsp	0.19166	0.18542741	6.8597	6.9694
	CONST	1.93735	N.A.	10.6741	N.A.
Between	lpcap	0.17131	0.1713115	2.2882	2.2825
	lpc	0.30163	0.3016278	7.1520	7.1343
	lemp	0.58559	0.5855900	9.6283	9.6045
	unemp	-0.00242	-0.0024207	-0.2297	-0.2291
	W*lgsp	-0.01082	-0.0108194	-0.4373	-0.4363
	CONST	1.70896	1.7089613	4.7423	4.7306
Random	lpcap	0.02286	0.01938433	0.9173	0.7823
	lpc	0.29376	0.29156673	13.9596	13.7588
	lemp	0.70864	0.71205474	26.4514	26.5722
	unemp	-0.00648	-0.00638432	-7.0525	-7.0221
	W*lgsp	0.03547	0.03645267	2.3946	2.4111
	CONST	1.90996	1.93191718	11.4865	11.6601

Table 3: Comparison of estimated coefficients and t statistics for spatial panel data against R.

7. Conclusions

The new **Panel Data Toolbox** covers a wide variety of panel data models in the organized environment provided by MATLAB. Estimation methods include fixed, between and random effects, as well as instrumental variables models and spatial models.

Numerical checks show the consistency of the results, as the estimated coefficients and t statistics are equal to those reported by Stata and R for panel and instrumental panel data methods. This positions the new toolbox as a valid self-contained alternative for panel data econometrics in MATLAB.

Future improvements aim at adding new econometric methods, including unbalanced and rotating panels, dynamic panel data models, and additional tests.

Acknowledgments

This research was supported by the Spanish Ministry of Science and Innovation under research grant (ECO2010-21643). Javier Barbero acknowledges financial support from the Spanish Ministry of Science and Innovation (AP2010-1401).

References

Anselin L (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Publisher, Dordrecht.

Anselin L (2010). "Thirty Years of Spatial Econometrics." *Papers in Regional Science*, **89**(1), 3–25. ISSN 1435-5957.

Baltagi B, Egger P, Pfaffermayr M (2006). "A Generalized Spatial Panel Data Model with Random Effects." *woking paper*, Center For Policy Research, Syracuse University.

Baltagi BH (1981). "Simultaneous Equations With Error Components." *Journal of Econometrics*, **17**(2), 189–200.

Baltagi BH (2008). *Econometric Analysis of Panel Data*. 4th edition. John Wiley & Sons Ltd, United Kingdom.

Baltagi BH, Griffin JM, Xiong W (2000). "To Pool Or Not To Pool: Homogeneous Versus Hetergeneous Estimations Applied to Cigarette Demand." *The Review of Economics and Statistics*, **82**(1), 117–126.

Baltagi BH, Levin D (1992). "Cigarette Taxation: Raising Revenues and Teducing Consumption." *Structural Change and Economic Dynamics*, **3**(2), 321–335.

Baltagi BH, Liu L (2009). "A Note on the Application of EC2SLS and EC3SLS Estimators in Panel Data Models." *Statistics & Probability Letters*, **79**(20), 2189–2192.

Baltagi BH, Liu L (2011). "Instrumental Variable Estimation of a Spatial Autoregressive Panel Model with Random Effects." *Economics Letters*, **111**(2), 135–137.

Croissant Y, Millo G (2008). "Panel Data Econometrics in R: The plm Package." *Journal of Statistical Software*, **27**(2), 1–43. ISSN 1548-7660. URL http://www.jstatsoft.org/v27/i02.

Elhorst JP (2003). "Unconditional Maximum Likelihood Estimation of Dynamic Models for Spatial Panels." *Research Report 03C27*, University of Groningen, Research Institute SOM (Systems, Organisations and Management).

Elhorst JP (2010). "Applied Spatial Econometrics: Raising the Bar." *Spatial Economic Analysis*, **5**(1), 9–28.

Elhorst JP (2011). "MATLAB Software to Estimate Spatial Panels." URL http://www.regroningen.nl/elhorst/software.shtml.

Greene WH (2012). *Econometric Analysis*. 7th edition. Prentice Hall, Upper Saddle River, New Jersey.

Hansen CB (2007). "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large." *Journal of Econometrics*, **141**(2), 597–620.

Hausman JA (1978). "Specification Tests in Econometrics." *Econometrica*, **46**(6), 1251–71.

Kapoor M, Kelejian HH, Prucha IR (2007). "Panel Data Models with Spatially Correlated Error Components." *Journal of Econometrics*, **140**(1), 97–130.

Kelejian HH, Prucha IR (1998). "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." *The Journal of Real Estate Finance and Economics*, **17**(1), 99–121.

Kelejian HH, Prucha IR (1999). "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review*, **40**(2), 509–33.

Kelejian HH, Prucha IR (2004). "Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations." *Journal of Econometrics*, **118**(1-2), 27–50.

Kelejian HH, Prucha IR (2007). "HAC Estimation in a Spatial Framework." *Journal of Econometrics*, **140**(1), 131–154.

Kelejian HH, Prucha IR (2010). "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances." *Journal of Econometrics*, **157**(1), 53–67.

Lee Lf (2003). "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances." *Econometric Reviews*, **22**(4), 307–335.

Lee Lf, Yu J (2010). "Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects." *Journal of Econometrics*, **154**(2), 165–185.

LeSage J, Pace RK (2009). *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.

Maddala GS, Mount TD (1973). "A Comparative Study of Alternative Estimators for Variance Components Models Used in Econometric Applications." *Journal of the American Statistical Association*, **68**(342), 324–328.

Millo G, Piras G (2012). "splm: Spatial Panel Data Models in R." *Journal of Statistical Software*, **47**(1), 1–38. ISSN 1548-7660. URL http://www.jstatsoft.org/v47/i01.

Munnell AH (1990). "Why has Productivity Growth Declined? Productivity and Public Investment." *New England Economic Review*, pp. 3–22.

Piras G (2010). "sphet: Spatial Models with Heteroskedastic Innovations in R." *Journal of Statistical Software*, **35**(1), 1–21. ISSN 1548-7660. URL http://www.jstatsoft.org/v35/i01.

Swamy PAVB, Arora SS (1972). "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models." *Econometrica*, **40**(2), pp. 261–275. ISSN 00129682.

The MathWorks, Inc (2013). *MATLAB — The Language of Technical Computing, Version R2013a (8.1)*. Natick, Massachusetts. URL http://www.mathworks.com/products/matlab/.

White H (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, **48**(4), 817–838.

Affiliation:

Inmaculada C.Álvarez, Javier Barbero, José L. Zofío
Department of Economics
Universidad Autónoma de Madrid
28049 Madrid, Spain
E-mail: inmaculada.alvarez@uam.es, javier.barbero@uam.es, jose.zofio@uam.es
URL: http://www.paneldatatoolbox.com/