Distributed-Lag Structural Equation Modelling with the R Package dlsem

Alessandro Magrini
Dep. Statistics, Computer Science, Applications
University of Florence, Italy
<magrini@disia.unifi.it>

`dlsem` version 1.5 – 23 November 2016

Contents

1 Introduction

Package `dlsem` implements inference functionalities for structural equation modelling with second-order polynomial and gamma lag shapes (DLSEM, [Magrini *et al.* (2016)]). DLSEM is an extension of structural equation modelling (SEM) where a second-order polynomial or a gamma lag shape is applied to each covariate in each regression model, in order to account for temporal delays in the dependence relationships among variables. In this vignette, theory on structural equation modelling with second-order polynomial and gamma lag shapes is provided in Section 2, then the practical use of `dlsem` is illustrated in Section 3 through a fictitious impact assessment problem.

2 Theory

Lagged instances of one or more quantitative covariates can be included in the classical linear regression model to to account for temporal delays in their influence on the response:

$$y_t = \beta_0 + \sum_{j=1}^{J} \sum_{l=0}^{L_j} \beta_{j,l}\ x_{j,t-l} + \epsilon_t \qquad \epsilon_t \sim \mathrm{N}(0, \sigma^2) \tag{1}$$

where y_t is the value of the response variable at time t and $x_{j,t-l}$ is the value of the j-th covariate at l time lags before t. The set $(\beta_{j,0}, \beta_{j,1}, \ldots, \beta_{j,L_j})$ is denoted as the *lag shape* of the j-th covariate and represents its effect on the response variable at different time lags.

Parameter estimation using ordinary least squares is inefficient because lagged instances of the same covariate are typically highly correlated. Also, the lag shape of a covariate is completely unrestricted, thus problems of interpretation may arise. Second-order polynomial and gamma lag shapes can be used to solve these drawbacks [Baltagi (2008), Chapter 6]. Package `dlsem` includes the endpoint-constrained quadratic lag shape:

$$\beta_{j,l} = \begin{cases} \theta_j \left[-\frac{4}{(b_j-a_j+2)^2} l^2 + \frac{4(a_j+b_j)}{(b_j-a_j+2)^2} l - \frac{4(a_j-1)(b_j+1)}{(b_j-a_j+2)^2} \right] & a_j \leq l \leq b_j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

the quadratic decreasing lag shape:

$$\beta_{j,l} = \begin{cases} \theta_j \frac{l^2-2b_j l+b_j^2}{(b_j-a_j)^2} & a_j \leq l \leq b_j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and the gamma lag shape:

$$\beta_{j,l} = \theta_j (l+1)^{\frac{\delta}{1-\delta}} \lambda_j^l \left[\left(\frac{\delta_j}{(\delta_j-1)\log(\lambda_j)} \right)^{\frac{\delta_j}{1-\delta_j}} \lambda_j^{\frac{\delta_j}{(\delta_j-1)\log(\lambda_j)}-1} \right]^{-1} \tag{4}$$
$$0 < \delta_j < 1 \qquad 0 < \lambda_j < 1.$$

The endpoint-constrained quadratic lag shape is zero for a lag $l \leq a_j - 1$ or $l \geq b_j + 1$, and symmetric with mode equal to θ_j at $(a_j + b_j)/2$. The quadratic decreasing lag shape decreases from value θ_j at lag a_j to value 0 at lag b_j according to a quadratic function. The gamma lag shape is positively skewed with mode equal to θ_j at $\frac{\delta_j}{(\delta_j-1)\log(\lambda_j)}$. Value a_j is denoted as the *gestation lag*, and value $b_j - a_j$ as the *lag width*. A static regression coefficient is obtained if $a_j = b_j = 0$. Since it is not expressed as a function of a_j and b_j, the gamma lag shape cannot reduce to a static regression coefficent, but values a_j and b_j can be computed through numerical approximation. A second-order polynomial or a gamma lag shape is monotonic in the sign, that is $\beta_{j,l}$ is either non-negative or non-positive for any j and l.

A linear regression model with second-order polynomial and gamma lag shapes is linear in parameters $\beta_0, \theta_1, \ldots, \theta_J$, provided that the values of $a_1, \ldots, a_J, b_1, \ldots, b_J$ are known. Thus, one can use ordinary least squares to estimate parameters $\beta_0, \theta_1, \ldots, \theta_J$ for several models with different values of $a_1, \ldots, a_J, b_1, \ldots, b_J$, and then select the one with the lowest Akaike Information Criterion[1].

In structural equation modelling (SEM), a linear regression model is applied to each variable and all linear regression models define an acyclic directed graph (DAG). In such DAG, variables are represented by nodes, a node receives a directed edge from another node if the variable represented by the latter is a covariate in the regression model of the variable represented by the former, and no directed cycles are present (see Figure 1). If a node receives a directed edge from another node in the DAG, the former is called child of the latter, and the latter is called parent of the former. A comprehensive review of SEM can be found in [Kline (2000)]. If the DAG has a causal interpretation, a causal effect is associated to each edge, directed path or couple of variables [Pearl (2012)]:

- the causal effect associated to each edge in the DAG is represented by the coefficient of the variable represented by the parent node in the regression model of the variable represented by the child node;
- the causal effect associated to a directed path is represented by the product of the causal effects associated to each edge in the path;

[1]Neither the response variable nor the covariates must contain a trend in order to obtain unbiased estimates [Granger and Newbold (1974)]. A reasonable procedure is to sequentially apply differentiation to all variables until the Dickey-Fuller test rejects the hypothesis of unit root for all of them.

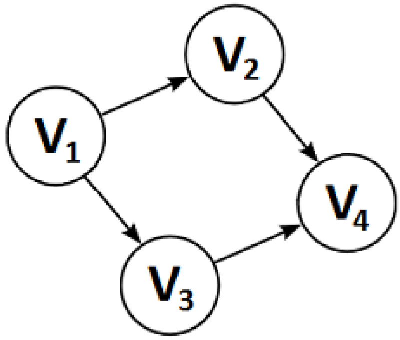

Figure 1: A directed acyclic graph for SEM. The regression model applied to variable V_1 has no covariates, the regression models applied to variables V_2 and V_3 have V_1 as covariate, the regression model applied to variable V_4 has V_2 and V_3 as covariates.

- the causal effect of a variable on another is represented by the sum of the causal effects associated to each directed path connecting the two variables.

Often, the causal effect of a variable on another is termed *overall* causal effect, the causal effect associated to a directed path made by a single edge is called *direct* effect, while the causal effects associated to the other directed paths are denoted as *indirect* effects.

In distributed-lag structural equation modelling (DLSEM), each regression model is enhanced by second-order polynomial and gamma lag shapes and the DAG does not explicitly include time lags, but, if an edge connects two variables, then there must be at least one time lag where the coefficient of the variable represented by the parent node in the regression model of the variable represented by the child node is non-zero. DLSEM can be employed to disentangle the causal effect of any variable to another at different time lags by extending the rules above:

- The causal effect associated to each edge in the DAG at lag k is represented by the coefficient at lag k of the variable represented by the parent node in the regression model of the variable represented by the child node.
- The causal effect associated to a directed path at lag k is computed as follows:
  1. denote the number of edges in the path as p;
  2. enumerate all the possible p-uples of lags, one lag for each of the p edges, such that their sum is equal to k;
  3. for each p-uple of lags:
     - for each lag in the p-uple, compute the coefficient associated to the corresponding edge at that lag;
     - compute the product of all these coefficients;
  4. sum all these products.
- The causal effect of a variable on another at lag k is represented by the sum of the causal effects at lag k associated to each directed path connecting the two variables.

A causal effect evaluated at a single lag is denoted as *instantaneous* causal effect. The *cumulative* causal effect at a prespecified lag, say k, is obtained by summing all the instantaneous causal effects for each lag up to k.

3 Distributed-lag structural equation modelling with dlsem

The practical use of package `dlsem` is illustrated through a fictitious impact assessment problem, aiming at testing whether the influence through time of the number job positions in industry (proxy of the industrial development) on the amount of greenhouse gas emissions (proxy of pollution) is direct and/or mediated by the amount of private consumption. The analysis will be conducted on the dataset `industry`, containing data for 10 immaginary regions in the period 1983-2015.

```
> data(industry)
> summary(industry)
```

```
     Region          Year        Population             GDP
 Min.   : 1.0   Min.   :1983   Min.   : 4763886   Min.   :  95430
 1st Qu.: 3.0   1st Qu.:1991   1st Qu.: 8320552   1st Qu.: 182835
 Median : 5.5   Median :1998   Median :25494251   Median : 461110
 Mean   : 5.5   Mean   :1998   Mean   :32404815   Mean   : 722088
 3rd Qu.: 8.0   3rd Qu.:2006   3rd Qu.:56325617   3rd Qu.:1272378
 Max.   :10.0   Max.   :2014   Max.   :78317539   Max.   :1901180
      Job               Consum          Pollution
 Min.   :   1380   Min.   : 31.96   Min.   : 5715
 1st Qu.:  48138   1st Qu.: 85.98   1st Qu.: 7906
 Median :  90386   Median :101.79   Median :23918
 Mean   : 254451   Mean   : 98.71   Mean   :32329
 3rd Qu.: 432675   3rd Qu.:114.30   3rd Qu.:45515
 Max.   :1666585   Max.   :173.53   Max.   :94038
```

3.1 The model code

The first step to perform DLSEM with `dlsem` is the specification of the model code encoding the DAG relating variables, together with assumptions and constraints on the lag shape for each variable. The DAG for the proposed problem is shown in Figure 2.

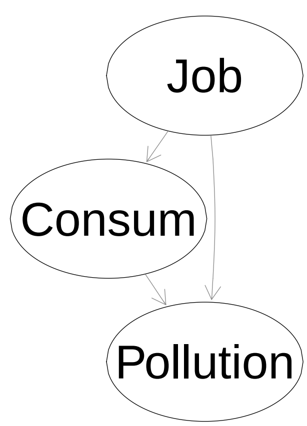

Figure 2: The DAG for the industrial development problem. 'Job': number of job positions in industry. 'Consum': private consumption index. 'Pollution': amount of greenhouse gas emissions.

The model code must be a list of formulas, one for each regression model. In each formula, the response and the covariates must be quantitative variables and operators `quec( )`, `qdec( )` and `gamma( )` can be employed to specify, respectively, an endpoint-constrained quadratic, a quadratic decreasing or a gamma lag shape. Operators `quec( )` and `qdec( )` have three arguments: the name of the variable to which the lag shape is applied, the minimum lag with a non-zero coefficient (a_j), and the maximum lag with a non-zero coefficient (b_j). Operator `gamma( )` has three arguments: the name of the variable to which the lag shape is applied, parameter δ_j and parameter λ_j. If none of these two operators is applied to a variable, it is assumed that the coefficient associated to that variable is 0 for time lags greater than 0 (no lag). The group factor and exogenous variables must not be specified in the model code (see Subsection 3.3). The regression model for variables with no covariates besides the group factor and exogenous variables can be omitted from the model code (here, we could omit the regression model for the number of job positions). In this illustration, an endpoint-constrained quadratic lag shape between 0 and 15 time lags is assumed for all variables:

```
> mycode <- list(
+   Job ~ 1,
+   Consum~quec(Job,0,15),
+   Pollution~quec(Job,0,15)+quec(Consum,0,15)
+   )
```

3.2 Control options

The second step to perform DLSEM with `dlsem` is the specification of control options. Control options must be a named list containing one or more among several components. The key component is adapt, a named vector of logical values where each value must refer to one response variable and indicates whether values a_j and b_j for each lag shape in the regression model of that variable must be selected on the basis of the best fit to data, instead of employing the ones specified in the model code. If adaption is requested for a regression model, three further components are taken into account: max.gestation, min.width, max.width and sign. Each of these three components is a named list, where each component of the list must refer to one response variable and must be a named vector including, respectively, the maximum gestation lag, the minimum lag width, the maximum lag width and the sign (either '+' for non-negative, or '-' for non-positive) of the coefficients of one or more covariates. In this illustration, adaptation of lag shapes is performed for all regression models with the following constraints: (i) maximum gestation lag of 3 years, (ii) minimum lag width of 5 years, (iii) maximum lag width of 15 years, (iv) all coefficients with non-negative sign

```
> mycontrol <- list(
+   adapt=c(Consum=T,Pollution=T),
+   max.gestation=list(Consum=c(Job=3),Pollution=c(Job=3,Consum=3)),
+   min.width=list(Consum=c(Job=5),Pollution=c(Job=5,Consum=5)),
+   max.width=list(Consum=c(Job=15),Pollution=c(Job=15,Consum=15)),
+   sign=list(Consum=c(Job="+"),Pollution=c(Job="+",Consum="+"))
+   )
```

3.3 Estimation

Once the model code and control options are specified, the structural model can be estimated from data using the command dlsem( ). The user can indicate a group factor to argument group and one or more exogenous variables to argument exogenous. By indicating the group factor, one intercept for each level of the group factor will be estimated in each regression model. By indicating exogenous variables, they will be included as non-lagged covariates in each regression model, in order to eliminate spurious effects due to differences between the levels of the group factor. Each exogenous variable can be either qualitative or quantitative and its coefficient in each regression model is 0 for time lags greater than 0 (no lag). Furthermore, the user can decide to perform any number of the following operations:

- differentiation until the hypothesis of unit root is rejected by the Dickey-Fuller test for all the quantitative variables (by setting argument uniroot.check to TRUE);
- apply the logarithmic transformation to all quantitative variables in order to interpret each coefficient as an elasticity (by setting argument log to TRUE).

In this illustration, the region is indicated as the group factor, while population and gross domestic product are indicated as exogenous variables. Also, we request differentiation until stationarity and logarithmic transformation for all quantitative variables:

```
> mod0 <- dlsem(mycode,group="Region",exogenous=c("Population","GDP"),
+   data=industry,control=mycontrol,uniroot.check=T,log=T)
```

```
Checking stationarity...
 Order 1 differentiation performed
Start estimation...
 Estimating regression model 1/3 (Job)
 Estimating regression model 2/3 (Consum)
 Estimating regression model 3/3 (Pollution)
Estimation completed
```

Before estimating the structural model, missing values for quantitative variables are imputed using the Expectation-Maximization algorithm [Dempster *et al.* (1977)]. After estimating the structural model, the user can display the DAG where each edge is coloured according to the sign of its causal effect (green for non-negative, red for non-positive). The result is shown in Figure 3: the group factor and exogenous variables are omitted from the DAG.

```
> plot(mod0)
```

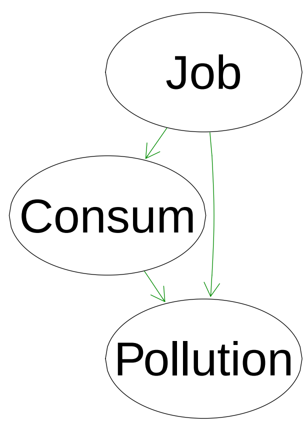

Figure 3: The DAG where each edge is coloured with respect to the sign of its causal effect. Green: non-negative causal effect. Red: non-positive causal effect. Grey: not statistically significant causal effect (does not apply in this example.

All edges result statistically significant, providing evidence that the influence of industrial development on pollution is both direct and mediated by private consumption.

The user can also request the summary of estimation:

```
> summary(mod0)
```

```
$Job

Call:
"Job ~ Region+Population+GDP"

Residuals:
    Min      1Q  Median      3Q     Max 
-5.9062 -0.3286 -0.0015  0.2867  5.1155 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
factor(Region)1  -0.057067   0.146386  -0.390   0.6969    
```

```
factor(Region)2  -0.031308   0.146290  -0.214   0.8307
factor(Region)3  -0.021608   0.146294  -0.148   0.8827
factor(Region)4  -0.161189   0.146382  -1.101   0.2717
factor(Region)5  -0.017219   0.146289  -0.118   0.9064
factor(Region)6  -0.027642   0.146277  -0.189   0.8502
factor(Region)7  -0.005684   0.146293  -0.039   0.9690
factor(Region)8  -0.046469   0.146304  -0.318   0.7510
factor(Region)9  -0.078167   0.146327  -0.534   0.5936
factor(Region)10 -0.046289   0.146356  -0.316   0.7520
Population       37.503956  22.489366   1.668   0.0964 .
GDP              -1.760618   1.981745  -0.888   0.3750
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8144 on 298 degrees of freedom
  (10 observations deleted due to missingness)
Multiple R-squared:  0.01926,        Adjusted R-squared:  -0.02023
F-statistic: 0.4877 on 12 and 298 DF,  p-value: 0.9214


$Consum

Call:
"Consum ~ Region+quec(Job,0,6)+Population+GDP"

Residuals:
      Min        1Q    Median        3Q       Max
-0.222080 -0.028044  0.000101  0.025323  0.303720

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
factor(Region)1  -4.158e-05  1.223e-02  -0.003  0.99729
factor(Region)2  -1.257e-02  1.221e-02  -1.030  0.30403
factor(Region)3   5.516e-03  1.220e-02   0.452  0.65164
factor(Region)4   8.464e-03  1.238e-02   0.684  0.49484
factor(Region)5   7.123e-03  1.220e-02   0.584  0.55994
factor(Region)6   1.761e-02  1.220e-02   1.443  0.15045
factor(Region)7   2.013e-02  1.221e-02   1.649  0.10054
factor(Region)8  -2.325e-02  1.222e-02  -1.902  0.05836 .
factor(Region)9   7.496e-03  1.222e-02   0.613  0.54035
factor(Region)10 -5.894e-03  1.222e-02  -0.482  0.62993
theta0_quec.Job   3.173e-02  1.533e-02   2.070  0.03954 *
Population       -4.839e+00  1.729e+00  -2.798  0.00557 **
GDP              -8.555e-01  1.553e-01  -5.510 9.35e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06098 on 237 degrees of freedom
  (70 observations deleted due to missingness)
Multiple R-squared:  0.1758,        Adjusted R-squared:  0.1306
F-statistic: 3.889 on 13 and 237 DF,  p-value: 1.017e-05


$Pollution

Call:
"Pollution ~ Region+quec(Job,1,11)+quec(Consum,1,6)+Population+GDP"

Residuals:
      Min        1Q    Median        3Q       Max
-0.127137 -0.019349 -0.001778  0.018852  0.107878

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
factor(Region)1     0.002500   0.007729   0.323 0.746698
```

```
factor(Region)2     0.012827   0.008611   1.490 0.138019
factor(Region)3    -0.002839   0.007850  -0.362 0.718000
factor(Region)4    -0.001595   0.008365  -0.191 0.849031
factor(Region)5    -0.004498   0.008174  -0.550 0.582806
factor(Region)6    -0.017220   0.008739  -1.970 0.050278 .
factor(Region)7    -0.020123   0.009491  -2.120 0.035310 *
factor(Region)8     0.022969   0.010492   2.189 0.029830 *
factor(Region)9    -0.005865   0.007925  -0.740 0.460140
factor(Region)10    0.007015   0.007845   0.894 0.372359
theta0_quec.Job     0.037419   0.013294   2.815 0.005408 **
theta0_quec.Consum  0.230538   0.063627   3.623 0.000375 ***
Population         -3.333542   1.604427  -2.078 0.039109 *
GDP                 0.181482   0.093399   1.943 0.053516 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03451 on 186 degrees of freedom
  (120 observations deleted due to missingness)
Multiple R-squared:  0.167,	Adjusted R-squared:  0.1043
F-statistic: 2.663 on 14 and 186 DF,  p-value: 0.001446
```

The summary of estimation returns estimates of parameters θ_j $(j = 1, \ldots, J)$. Instead, the command edgeCoeff( ) can be used to obtain estimates and confidence intervals of coefficients at the relevant time lags $\beta_{j,l}$ $(j = 1, \ldots, J; l = 0, 1, \ldots)$:

```
> edgeCoeff(mod0)
$`0`
                         2.5%        50%      97.5%
Consum~Job       0.0007376091 0.01388267 0.02702774
Pollution~Job    0.0000000000 0.00000000 0.00000000
Pollution~Consum 0.0000000000 0.00000000 0.00000000

$`1`
                        2.5%        50%      97.5%
Consum~Job       0.001264473 0.02379887 0.04633326
Pollution~Job    0.003472057 0.01143358 0.01939511
Pollution~Consum 0.051835751 0.11291665 0.17399755

$`2`
                        2.5%        50%      97.5%
Consum~Job       0.001580591 0.02974858 0.05791658
Pollution~Job    0.006312831 0.02078834 0.03526384
Pollution~Consum 0.086392919 0.18819441 0.28999591

$`3`
                        2.5%        50%      97.5%
Consum~Job       0.001685964 0.03173182 0.06177768
Pollution~Job    0.008522322 0.02806425 0.04760618
Pollution~Consum 0.103671503 0.22583330 0.34799509

$`4`
                        2.5%        50%      97.5%
Consum~Job       0.001580591 0.02974858 0.05791658
Pollution~Job    0.010100529 0.03326134 0.05642214
Pollution~Consum 0.103671503 0.22583330 0.34799509

$`5`
                        2.5%        50%      97.5%
Consum~Job       0.001264473 0.02379887 0.04633326
Pollution~Job    0.011047454 0.03637959 0.06171172
Pollution~Consum 0.086392919 0.18819441 0.28999591

$`6`
                         2.5%        50%      97.5%
```

```
Consum~Job       0.0007376091 0.01388267 0.02702774
Pollution~Job    0.0113630955 0.03741900 0.06347491
Pollution~Consum 0.0518357514 0.11291665 0.17399755

$`7`
                       2.5%        50%      97.5%
Consum~Job       0.00000000 0.00000000 0.00000000
Pollution~Job    0.01104745 0.03637959 0.06171172
Pollution~Consum 0.00000000 0.00000000 0.00000000

$`8`
                       2.5%        50%      97.5%
Consum~Job       0.00000000 0.00000000 0.00000000
Pollution~Job    0.01010053 0.03326134 0.05642214
Pollution~Consum 0.00000000 0.00000000 0.00000000

$`9`
                        2.5%        50%      97.5%
Consum~Job       0.000000000 0.00000000 0.00000000
Pollution~Job    0.008522322 0.02806425 0.04760618
Pollution~Consum 0.000000000 0.00000000 0.00000000

$`10`
                        2.5%        50%      97.5%
Consum~Job       0.000000000 0.00000000 0.00000000
Pollution~Job    0.006312831 0.02078834 0.03526384
Pollution~Consum 0.000000000 0.00000000 0.00000000

$`11`
                        2.5%        50%      97.5%
Consum~Job       0.000000000 0.00000000 0.00000000
Pollution~Job    0.003472057 0.01143358 0.01939511
Pollution~Consum 0.000000000 0.00000000 0.00000000
```

3.4 Disentanglement of causal effects

Causal effects can be computed using the command causalEff( ). The user must specify one or more starting variables (argument from) and the ending variable (argument to). Optionally, specific time lags at which causal effects must be computed can be provided to argument lag, otherwise all the relevant ones are considered. Also, the user can choose whether instantaneous (argument cumul set to FALSE, the default) or cumulative (argument cumul set to TRUE) causal effects must be returned. Here, the cumulative causal effect of the number of job positions on the amount of greenhouse gas emissions is requested at time lags 0, 5, 10, 15 and 20:

```
> causalEff(mod0,from="Job",to="Pollution",lag=seq(0,20,by=5),cumul=T)

$`Job*Consum*Pollution`
         2.5%        50%      97.5%
0  0.00000000 0.00000000 0.00000000
5  0.02352582 0.05956822 0.09561063
10 0.07442555 0.16870198 0.26297841
15 0.07533235 0.17556950 0.27580664
20 0.07533235 0.17556950 0.27580664

$`Job*Pollution`
         2.5%       50%     97.5%
0  0.00000000 0.0000000 0.0000000
5  0.03945519 0.1299271 0.2203990
10 0.08680142 0.2858396 0.4848778
15 0.09027348 0.2972732 0.5042729
20 0.09027348 0.2972732 0.5042729

$overall
```

```
         2.5%       50%     97.5%
0  0.00000000 0.0000000 0.0000000
5  0.06298101 0.1894953 0.3160096
10 0.16122698 0.4545416 0.7478562
15 0.16560583 0.4728427 0.7800796
20 0.16560583 0.4728427 0.7800796
```

The output of command causalEff is a list of matrices, each containing estimates and confidence intervals of the causal effect associated to each path connecting the starting variables to the ending variable at the requested time lags. Also, estimates and confidence intervals of the overall causal effect is shown in the component named overall.

Since the logarithmic trasformation was applied to all quantitative variables, causal effects above are interpreted as elasticities, that is, for a 1% of job positions more, greenhouse gas emissions are expected to grow by 0.47% after 20 years. Actually, the effect ends before 20 years, as the cumulative causal effects after 15 and20 years are equal. The time lag up to which the effect is non-zero can be found by running command causalEff without providing a value to argument lag:

```
> causalEff(mod0,from="Job",to="Pollution",cumul=T)
```

```
$`Job*Consum*Pollution`
            2.5%         50%       97.5%
0   0.0000000000 0.000000000 0.000000000
1  -0.0001902551 0.001567585 0.003325425
2   0.0009067999 0.006867514 0.012828229
3   0.0046845282 0.017840608 0.030996688
4   0.0121660315 0.035531922 0.058897813
5   0.0235258156 0.059568222 0.095610629
6   0.0376661767 0.087784749 0.137903321
7   0.0518065378 0.116001275 0.180196013
8   0.0631663218 0.140037575 0.216908829
9   0.0706478251 0.157728890 0.244809954
10  0.0744255535 0.168701983 0.262978413
11  0.0755226084 0.174001913 0.272481217
12  0.0753323533 0.175569498 0.275806642
13  0.0753323533 0.175569498 0.275806642

$`Job*Pollution`
          2.5%        50%      97.5%
0  0.000000000 0.00000000 0.00000000
1  0.003472057 0.01143358 0.01939511
2  0.009784888 0.03222192 0.05465895
3  0.018307209 0.06028617 0.10226514
4  0.028407739 0.09354751 0.15868728
5  0.039455193 0.12992710 0.22039900
6  0.050818288 0.16734610 0.28387391
7  0.061865742 0.20372569 0.34558563
8  0.071966271 0.23698702 0.40200777
9  0.080488593 0.26505128 0.44961396
10 0.086801424 0.28583961 0.48487780
11 0.090273481 0.29727320 0.50427291
12 0.090273481 0.29727320 0.50427291
13 0.090273481 0.29727320 0.50427291

$overall
          2.5%        50%      97.5%
0  0.000000000 0.00000000 0.00000000
1  0.003281802 0.01300117 0.02272054
2  0.010691688 0.03908943 0.06748718
3  0.022991738 0.07812678 0.13326182
4  0.040573770 0.12907943 0.21758509
5  0.062981008 0.18949532 0.31600963
6  0.088484465 0.25513085 0.42177723
7  0.113672280 0.31972696 0.52578164
```

```
8  0.135132593 0.37702460 0.61891660
9  0.151136418 0.42278016 0.69442391
10 0.161226977 0.45454159 0.74785621
11 0.165796089 0.47127511 0.77675413
12 0.165605834 0.47284269 0.78007955
13 0.165605834 0.47284269 0.78007955
```

The estimated lag shape associated to a path or to an overall causal effect can be displayed using the command lagPlot( ). For instance, we can display the lag shape associated to each path connecting the number of job positions to the amount of greenhouse gas emissions:

```
> lagPlot(mod0,path="Job*Pollution")
> lagPlot(mod0,path="Job*Consum*Pollution")
```

or the lag shape associated to the overall causal effect of the number of job positions on the amount of greenhouse gas emissions:

```
> lagPlot(mod0,from="Job",to="Pollution")
```

The resulting graphics are shown in Figure 4.

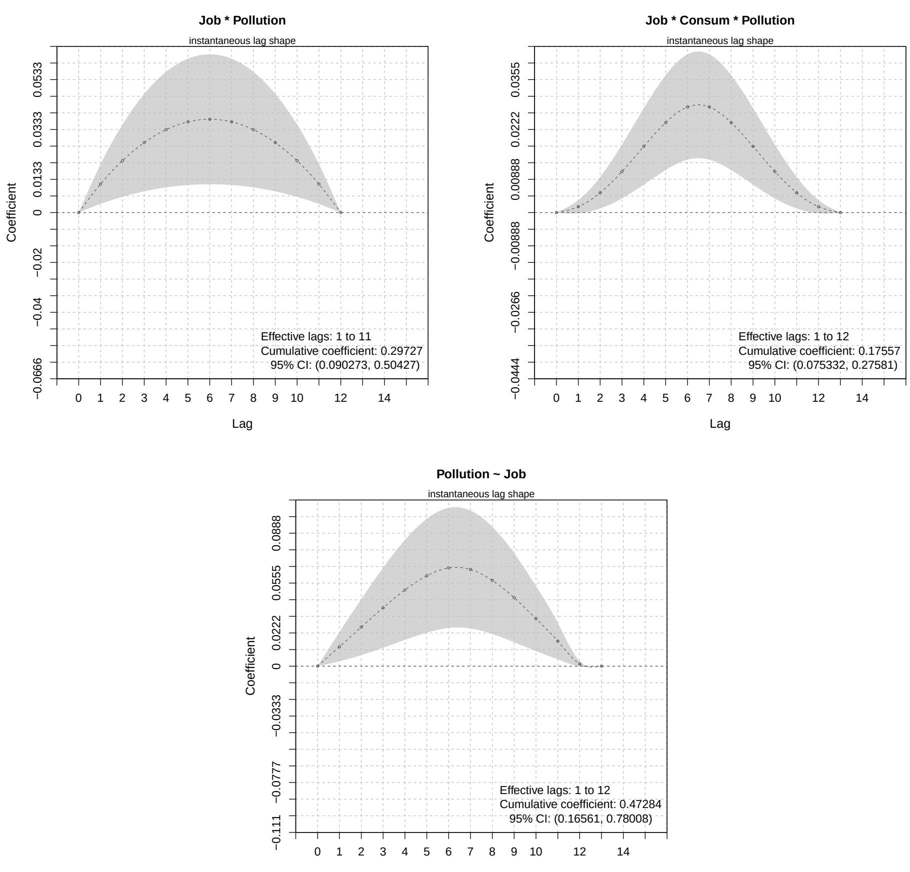

Figure 4: The estimated lag shape associated to each path connecting the number of job positions to the amount of greenhouse gas emissions (upper panels) and to the overall causal effect (lower panel). 95% confidence intervals are shown in grey.

References

[Baltagi (2008)] B. H. Baltagi (2008). Econometrics. Springer Verlag, 4th edition, Berlin, DE.

[Dempster *et al.* (1977)] A. P. Dempster, N. M. Laird, and D. B. Rubin (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, Series B, 39(1): 1-38.

[Granger and Newbold (1974)] C. W. J. Granger, and P. Newbold (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, 2(2), 111-120.

[Kline (2000)] R. B. Kline (2010). Principles and Practice of Structural Equation Modelling. Guilford Press, 3rd edition, New York, US-NY.

[Magrini *et al.* (2016)] A. Magrini, F. Bartolini, A. Coli, and B. Pacini (2016). Distributed-Lag Structural Equation Modelling: An Application to Impact Assessment of Research Activity on European Agriculture. *Proceedings of the 48th Meeting of the Italian Statistical Society*, 8-10 June 2016, Salerno, IT.

[Pearl (2012)] J. Pearl (2012). The Causal Foundations of Structural Equation Modelling. In: R. H. Hoyle (ed.), Handbook of Structural Equation Modelling, Chapter 5. Guilford Press, New York, US-NY.