**Additional file 3. Section Mathematics/formula**

**Multilayer Feed-forward Neural Networks model:** Following the neural network in Additional file 2, the inputs into the *j*-th hidden neuron are linearly combined to give

$$θ\_{j}= β\_{j}+ \sum\_{i=1}^{4}w\_{i,j}x\_{i}$$

Where $β\_{j}$ and the “weights” $w\_{i,j}$ are learned from the data and $x\_{i}$ is the vector of exogenous variables (i.e. the regressors). In the hidden layer, the result is transformed using a non-linear function, such as a sigmoid/logistic function

$s\left(θ\right)= \frac{1}{1+ e^{-θ}}$ ,

to give the input for the next layer, in order to reduce the effect of outliers.

When managing time series, lagged values of the time series can be included in the neural network as inputs. In this paper, we implemented a neural network auto-regression (i.e. a neural network that depends on exogenous values and dependent variable lagged values), NNAR(*p*,*k*), where *p* denotes the lagged inputs and *k* is equal to the number of nodes in the hidden layer. The nnetar() function from the package forecast in R allows to fit a feed-forward neural network, specifying the number of lags and hidden nodes. Since we rely on a single lag of the dependent variable and two nodes in the hidden layer, our predictive model is specified as a NNAR(1,2).

**Fixed effect regression model:** The fixed effect regression model can be defined as follow:

$$MR\_{i,t}= η\_{i}+ γHE\_{i,t}+ δX\_{i,t}+ θT+ ε\_{i,t}$$

where $γ$ is the vector of regression coefficients for expenditure variables (HE), $δ$ is the vector of regression coefficients for the control variables X (which vary across the models), $θ$ is the regression coefficient for the time trend T, while $η\_{i}$ is the unknown intercept for each of the 20 Italian Regions and $ε\_{i,t}$ is the error term.