CS-BIGS 4(2): 83-92

http://legacy.bentley.edu/csbigs/documents/bouveyron.pdf

Adaptive Regression Models for Modeling and Understanding Evolving Populations

Charles Bouveyron
Université Paris I Panthéon-Sorbonne, France

Patrice Gaubert
Université Paris 12, France

Julien Jacques
Université Lille I, France

When regression analysis is carried out with a prediction purpose, an evolution in the modeled phenomenon between the training and the prediction stages forces the statistician to start a new analysis. Similarly, when regression aims to explain the modeled phenomenon, a new regression model must be estimated whenever the phenomenon or its study conditions change. In this paper we show how the previous regression analysis can be used for the estimation of the regression model in the new situation, saving a new expensive data collection effort. Two case studies are considered: first a regression model of house prices versus house and household features is adapted from a city in the Southern United States (Birmingham, AL) to another city of the United States west coast (San Jose, CA). In the second case the link between CO_2 emissions and gross national product in 1999 is analyzed by using a previous analysis dating from 1980.

1. Introduction

In Economics as in many other fields, regression analysis is concerned with both the prediction of future phenomena and the interpretation of the data. In prediction, one of the main assumptions is the absence of evolution in the modeled phenomenon between the training and the prediction stages. If the assumption does not hold, a new regression model must be estimated independently from the previous analysis. For the same reasons, when the goal of the regression analysis is the interpretation of a phenomenon, studies of a same phenomenon but in different situations (at different periods of time, in different geographical places, etc.) are generally independently carried out.

In this work, it is shown how a regression model, used in order to predict or to explain a phenomenon in a given situation, can be efficiently adapted to a new situation. For this, adaptive models for linear regression (Bouveyron and Jacques 2003) and for mixture of regressions (Bouveyron and Jacques 2010) are considered.

In our first analysis, the goal is to predict house values in the city of San Jose (California, West coast) from several features such as housing units characteristics or socio-economic information about the households that occupy those units. We will see that using a regression model previously built for the city of Birmingham (Alabama, South) with the same variables can save a new expensive collection of data in the city of San Jose.

In the second study, a regression model of CO2 emissions in terms of the gross national product of countries is used for the explanation of the link between these two indicators. As in the previous study, we will see that data from 1980 and especially the regression model on these data can be useful for the estimation of a regression model on the 1999 data. Moreover, the exhibited link between the two regression models is informative and allows to explain the different evolutions of the economic policies of the considered countries.

The paper is organized as follows: Section 2 presents the two datasets whereas Section 3 briefly reviews the methodology. Results are then analyzed and discussed in Section 4.

2. The data

In this work, two datasets with evolving populations will be studied. This section briefly presents both datasets.

2.1. The American Housing Survey dataset

The first dataset which will be used in this study is the 1984 American Housing Survey (AHS) dataset. This is a statistical survey funded by the United States Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau. The AHS survey is the largest regular national housing sample survey in the United States; it aims to provide each year an overview of housing conditions in 11 U.S. metropolitan areas. This study focuses on two particular metropolitan areas: the cities of Birmingham, Alabama (South) and of San Jose, California (West coast). Fourteen relevant features have been selected among all available features for modeling the housing market of Birmingham. The dataset contains information on the number and characteristics of U.S. housing units as well as the households that occupy those units. The selected features for the study include the number of units in the property (NUNITS), the number of rooms (ROOMS), bedrooms (BEDRMS) and bathrooms (BATHS), the monthly housing cost (ZSMHC), the annual unit maintenance cost (CSTMNT), the monthly electricity cost (AMTE), the number of cars of the household (CARS), the unit area (UNITSF), the annual salary of the tenant (SAL1) and of the household (ZINC) and the number of persons in the household (PER). Finally, based on these 14 features, the response variable to predict is

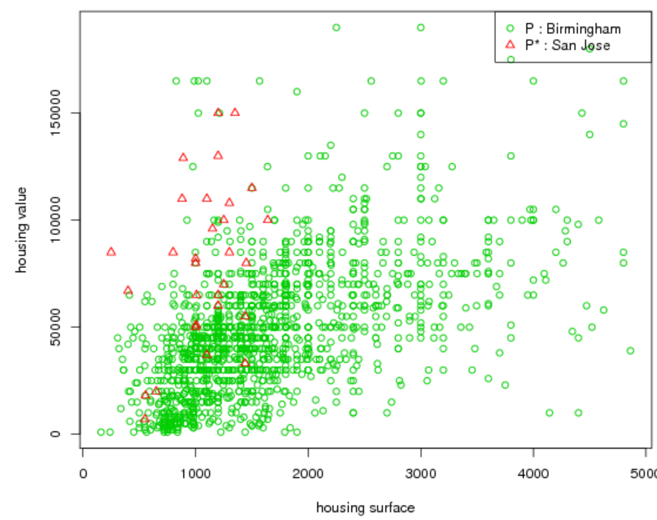

Figure 1. Housing value vs. area for Birmingham (AL, USA) and San Jose (CA,USA)

the value of the housing. The difference between Birmingham and of San Jose is illustrated by Figure 1 which presents the value of the houses in terms of their areas

In the present work we will see how a regression model of the house value estimated for the city of Birmingham can be adapted to the prediction of the houses values in San Jose.

2.2. The CO_2-GNP dataset

The economic aspects of the diffusion of greenhouse gases and their impact on the environment play an important role on the economies of countries, and their analysis has attracted a strong interest in the last twenty years (Barker 1991, Grubb and Ha-Duong 1997). As pointed out by Hurn et al. (2003), the study of such data could be particularly useful for countries with low GNP in order to clarify which development path they are embarking into.

The objectives of this study are to investigate the relationship between gross national product (GNP) and carbon dioxide gas (CO_2) emissions to contribute to current debates about emission projections. This study also aims to determine typical economic policies of countries regarding the environment. To this end, the second dataset studied in the present paper contains the CO_2 emissions per capita and the gross national product per capita for 111 countries in 1980 and 1999. The sources of the data are *The official United Nations site for the Millennium Development Goals Indicators and the World Development Indicators of the World Bank.* Figure 2 plots per capita for 111 countries, in 1980 (left) and 1999

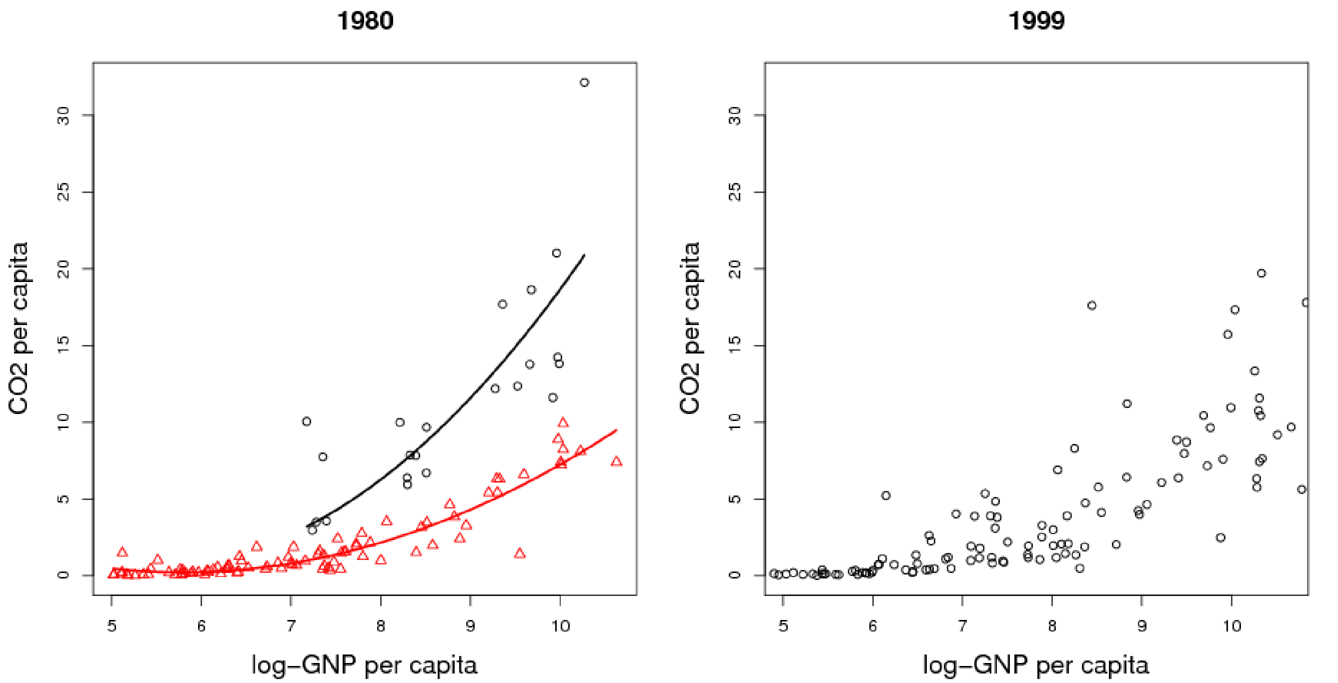

Figure 2. Emissions of CO_2 per capita versus GNP per capita in 1980 (left) and 1999 (right). In the left panel, circles indicate group 1 and triangles group 2.

CO_2 emissions per capita vs. the logarithm of GNP (right). The two groups of countries in the left panel of Figure 2 are discussed later in the paper.

We will see in this paper how the use of the 1980 data can be helpful in the analysis of 1999 data, by improving the quality of the regression models used to explain the relationship between the gross national product and the CO_2 emissions. Moreover, our analysis will shed light on the evolution of this relationship from 1980 to 1999, and will explain the economic political choices of particular countries.

3. Adaptive regression models

In this paper, the adaptive regression models proposed in Bouveyron and Jacques (2003 and 2010) will be used to analyze and understand the population evolution of the two datasets presented in the previous section. This section briefly reviews these adaptive regression models.

3.1. *Adaptive linear models*

The general setting of regression analysis is to identify a relationship (the regression model) between a response variable and one or several explanatory variables. Adaptive linear models have been defined in order to adapt an existing regression model to a new situation in which the variables are identical, but with a possibly different probability density and a relationship between response and explanatory variables which could have changed.

Linear models for regression In regression analysis, the data $S=\{(x_1,y_1),\ldots,(x_n,y_n)\}$ which arise from a population P, are assumed to be independent and identically distributed samples from an unknown distribution, where $x=(x^{(1)},\ldots,x^{(p)}) \in$ Rp and Y $\in$ R. In many regression situations, Y is considered as a stochastic variable and x as a deterministic one. A general data modeling problem is to identify the relationship between the explanatory variable x (or covariate) and the response variable Y (or dependent variable). Both standard parametric and non-parametric regression approaches start with the following model:

$$Y = f(x,\beta)+\varepsilon \,, \tag{1}$$

with $\varepsilon \sim N(0,\sigma^2)$ and where β is a vector of real-valued regression parameters.

The most common model is the linear form:

$$f(x,\beta)=\beta_0+\sum_{i=1}^{d}\beta_i\psi_i(x)\,, \tag{2}$$

with $\beta = (\beta_0, \beta_1, \ldots, \beta_d) \in$ Rd+1 is a vector of regression parameters, and $(\psi_i)_{1 \le i \le d}$ is a basis of regression functions. In particular, usual linear regression occurs when d=p and $\psi_i(x) = x^{(i)}$.

How to adapt a regression model to another population? Let us assume that the estimation of the regression function f has been obtained in a preliminary study by using the sample S, and that a new regression model must be adjusted to a new sample S*= $\{(x_1^*, y_1^*), \ldots, (x_n^*, y_n^*)\}$, measured on the same variables but arising from another population P* (n* is generally assumed to be small). The new regression model on P* can be written:

$$Y \mid x^* \sim N(f(x^*, \beta^*), \sigma^2)$$

with

$$f(x^*, \beta^*) = \beta_0^* + \sum_{i=1}^{d} \beta_i^* \psi_i(x^*).$$

Let us now specify the focus of adaptive linear models by making the following assumptions. First, the variables (Y, x) and (Y*, x *) are assumed to be the same but measured on two different populations. Second, the size n* of the observation sample S*= $(y_i^*, x_i^*)_{i=1,n^*}$ of population P* is assumed to be small compared to the number of observations from the reference population P. Otherwise, the mixture regression model could be estimated directly without using the training population.

We consider the following transformation model between both regression functions for modeling the link between both populations:

$$f(x^*, \beta^*) = \phi(f(x, \beta)) \tag{3}$$

Since the transformation model (3) proposed in the previous section is a very general model, we have to make additional assumptions on it. We propose to assume that the transformation function ϕ has the following form:

$$\phi(f(x, \beta)) = f(x, \lambda\beta)$$

with $\lambda \in$ Rd+1. This transformation can be also written in terms of the regression parameters of both models as follows:

$$\beta_i^* = \lambda_i \beta_i \quad \forall i = 1, \ldots, d, \tag{4}$$

with $\lambda_i \in$ R. We note that the regression functions ψ_i are assumed to be the same for both regression models, which is natural since the variables are identical in both populations.

A family of transformation models Since the number of parameters to estimate for the transformation (4) is equal to (d+1), learning this transformation model is equivalent to learning a new regression model from the sample S*. It is therefore necessary to reduce the number of free parameters and that can be done by imposing constraints on the transformation parameters λ_i . A family of seven transformation models is then considered, thereafter referred to as Adaptive Linear Models, from the most complex model (hereafter M0) to the simplest one (hereafter M6):

- Model M0: $\beta_0^* = \lambda_0 \beta_0$ and $\beta_i^* = \lambda_i \beta_i$ for i=1,…,d. This model is the most complex model of transformation between the two populations P and P*, and is equivalent to learning a new regression model from the sample S*.

- Model M1: $\beta_0^* = \beta_0$ and $\beta_i^* = \lambda_i \beta_i$ for i=1,...,d. This transformation model assumes that both regression models have the same intercept β_0.

- Model M2: $\beta_0^* = \lambda_0 \beta_0$ and $\beta_i^* = \lambda \beta_i$ for i=1,...,d. This transformation model assumes that the intercept of the two regression models differ by the multiple scalar λ_0 and that all the other regression parameters differ by the same multiple scalar λ .

- Model M3: $\beta_0^* = \lambda \beta_0$ and $\beta_i^* = \lambda \beta_i$ for i=1,...,d. This transformation model assumes that all the regression parameters of both regression models differ by the same multiple scalar λ .

- Model M4: $\beta_0^* = \beta_0$ and $\beta_i^* = \lambda \beta_i$ for i=1,...,d. This transformation model assumes that both regression models have the same intercept β_0 and all the other regression parameters differ by the same multiple scalar λ .

- Model M5: $\beta_0^* = \lambda_0 \beta_0$ and $\beta_i^* = \beta_i$ for i=1,…,d. This transformation model assumes that both regression models have the same parameters except the intercept.

Table 1. Complexity (number of parameters) of the transformation models. We recall that the models M0 and M6 correspond respectively to OLS on P* and OLS on P.

Model	M_0	M_1	M_2	M_3	M_4	M_5	M_6
Parameters numbers	d+1	d	2	1	1	1	0

- Model M6: $\beta_0^* = \beta_0$ and $\beta_i^* = \beta_i$ for i=1, ...,d. This model assume that both populations P and P* have the same behavior.

The numbers of parameters to estimate for these transformation models are presented in Table 1. Note that it is possible to consider intermediate models, by imposing specific constraints on some parameters λ_i for given i ∈ {1,…,d}. The practitioner could use domain knowledge to introduce some intermediate models which are useful for the application at hand.

Estimation procedure and model selection The estimation procedure consists of two main steps corresponding to the estimation of the regression parameters on the population P and the estimation of the transformation parameters using samples of the population P*. The natural method for the first estimation step is to use the ordinary least squares (OLS) procedure which aims to minimize the error sum of squares. Once the regression parameters of population P have been learned, the parameters of the transformation models can also be estimated by minimizing least squares error using the OLS procedure. However, for the sake of brevity, the corresponding estimators are not presented in this paper. Finally, the cross validation PRESS criterion (Allen 2974) is used in order to select the most appropriate model for the data among the seven Adaptive Linear Models.

3.2. *Adaptive mixture models*

As an alternative to linear models for modeling complex systems, finite mixtures of regressions are a popular approach, introduced in Goldfeld and Quandt (1973) as switching regression models. In particular, these models are often used in Economics for modeling phenomena with different phases. They assume that the dependent variable Y ∈ R can be linked to a covariate x = $\left(1, x^{(1)}, \ldots, x^{(p)}\right)$ ∈ Rp+1 by one of K possible regression models:

$$Y = x^t \beta_k + \sigma_k \varepsilon, \; k = 1, \ldots, K \tag{5}$$

where ε~N(0,1), $\beta_k = \left(\beta_{k0}, \ldots, \beta_{kp}\right) \in \{\beta_1, \ldots, \beta_K\}$ is the regression parameter vector in Rp+1 and $\sigma_k^2 \in \{\sigma_1^2, \ldots, \sigma_K^2\}$ is the residual variance. The conditional density distribution of Y given x is therefore:

$$p(y|x) = \sum_{k=1}^{K} \pi_k \phi(y|x^t \beta_k, \sigma_k^2), \tag{6}$$

where the $\pi_1, \ldots, \pi_K$ are the mixing proportions (with the classical constraint $\sum_{i=1}^{K} \pi_k = 1$), and $\phi(.|x^t \beta_k, \sigma_k^2)$ is the Gaussian density parameterized by its mean $x^t \beta_k$ and variance σ_k^2. In the same way as for adaptive linear models, the new population P* for which we want to predict Y is assumed to be different from the training population P. The mixture regression model for P* can be written as follows:

$$Y^* = x^{*t} \beta_k^* + \sigma_k^* \varepsilon^*$$

$$p(y^*|x^*) = \sum \pi_k^* \phi(y^*|x^{*t} \beta_k^*, \sigma_k^{*2}) \tag{7}$$

with ε^* ~N(0,1), $\beta_k^* \in \{\beta_1^*, \ldots, \beta_{K^*}^*\}$ and $\sigma_k^* \in \{\sigma_1^*, \ldots, \sigma_{K^*}^*\}$. In the addition to the assumptions made in the previous section, since both populations have the same nature, each mixture is assumed to have the same number of components (K*=K). Under these assumptions, the goal is then to predict Y* for some new x^* by using both samples S=(y_i, x i)i=1,n and S*.

A family of transformation models Following the strategy of the linear case, the following general transformation model is considered:

$$\beta_k^* = \Lambda_k \beta_k, \tag{8}$$

where $\Lambda_k = diag(\lambda_{k0}, \lambda_{k1}, \ldots, \lambda_{kp})$

σ_k^* is free,

where $diag(\lambda_{k0}, \lambda_{k1}, \ldots, \lambda_{kp})$ is the diagonal matrix containing ($\lambda_{k0}, \lambda_{k1}, \ldots, \lambda_{kp}$) on its diagonal completed by zeros. The family of parsimonious models is defined by imposing some constraints on Λ_k :

- MM1 assumes that both populations are the same population: Λ_k =Id is the identity matrix,

- MM2 assumes that the link between populations is independent of the covariates and mixture components:

  - MM2a : $\lambda_{k0} = 1$, $\lambda_{k0} = 1$ and $\sigma_k^* = \lambda\sigma_k$ $\forall 1 \le j \le p$,
  - MM2b : $\Lambda_k = \lambda I_d$, $\lambda_{kj} = 1$ and $\sigma_k^* = \sigma_k$ $\forall 1 \le j \le p$,
  - MM2c : $\Lambda_k = \lambda$ Id and $\sigma_k^* = \lambda\sigma_k$,
  - MM2d : $\lambda_{k0} = \lambda_0$, $\lambda_{kj} = \lambda_1$ and $\sigma_k^* = \lambda_1\sigma_k$ $\forall 1 \le j \le p$,

- MM3 assumes that the link between populations is independent of the covariates:

  - MM3a : $\lambda_{k0} = 1$, $\lambda_{kj} = \lambda_k$ and $\sigma_k^* = \lambda_k\sigma_k$ $\forall 1 \le j \le p$,
  - MM3b : $\lambda_{k0} = \lambda_k$, $\lambda_{kj} = 1$ and $\sigma_k^* = \sigma_k$ $\forall 1 \le j \le p$,
  - MM3c : $\Lambda_k = \lambda_k$ Id and $\sigma_k^* = \lambda_k\sigma_k$,
  - MM3d : $\lambda_{k0} = \lambda_{k0}$, $\lambda_{kj} = \lambda_{k1}$ and $\sigma_k^* = \lambda_{k1}\sigma_k$ $\forall 1 \le j \le p$,

- MM4 assumes that the link between populations is independent of the mixture components:

  - MM4a : $\lambda_{k0} = 1$, $\lambda_{kj} = \lambda_j$ $\forall 1 \le j \le p$,
  - MM4b : $\Lambda_k = \Lambda$, $\lambda_{kj} = 1$ with Λ a diagonal matrix,

- MM5 assumes that Λ_k is unconstrained, which leads to estimating the mixture regression model for P* by using only S*.

Moreover, the mixing proportions are allowed to be the same in each population or to be different between both populations P and P*. In the latter case, they consequently have to be estimated using the sample S*. Corresponding notations for the models are respectively

MM. and pMM.. Table 2 gives the number of parameters to estimate for each model. If the mixing proportions are different from P to P*, K-1 parameters to estimate must be added to these values.

Estimation procedure and model selection As before, the estimation procedure consists of two steps. The first step consists in estimating model parameters for the reference population P whereas the second one focuses on the estimation of the link parameters. The estimation of the mixture regression parameters β_k^* is performed afterwards by plug-in. Conversely to the case of linear models, parameter estimation cannot be conducted with the standard OLS procedure and the estimation has to be carried out by maximum likelihood using a missing data approach via the EM algorithm (Dempster, Laird and Rubin 1977). Finally, in order to select among the transformation models previously defined the most appropriate model of transformation between the populations P and P*, we propose to use the PRESS criterion (Allen 1974) or the Bayesian Information Criterion (BIC, Schwarz 1978).

Table 2. Number of parameters to estimate for each model of the proposed family.

Model	MM_1	MM_{2a-c}	MM_{2d}	MM_{3a-c}
Param.	0	1	2	K
Model	MM_{3d}	MM_{4a}	MM_{4b}	MM_5
Param.	2K	p+K	p+K+1	K(p+2)

4. Experimental results

4.1 The housing market data

Experimental setup A semi-log regression model for the Birmingham housing was learned using all 1541 available observations and the 7 adaptive linear models were then used to transfer the Birmingham regression model to the San Jose housing market. In order to evaluate the ability of the adaptive linear models to transfer the Birmingham knowledge to San Jose in different situations, the experiment protocol was applied for different sizes of San Jose samples ranging from 5 to 921 observations. For each dataset size, the San Jose samples were randomly selected among all available samples and the experiment was repeated 50 times for averaging the results. For each adaptive linear model, the PRESS criterion and the MSE were computed, by using the selected sample for PRESS and the whole San Jose dataset for MSE.

Experimental results Figure 3 displays the logarithm of the MSE for the different adaptive linear models in terms of the size of the used San Jose samples. Similarly, Figure 4 displays the logarithm of the PRESS criterion. First Figure 3 indicates that model M6, which corresponds to the Birmingham model, is actually not adapted for modeling the San Jose housing market since it does not yield a satisfactory MSE value. We notice that the curve corresponding to the MSE of the model M6 is constant since the regression model has been learned on the Birmingham data and consequently does not depend on the size of the San Jose's dataset selected for learning. Secondly, the model M0, which is equivalent to OLS on

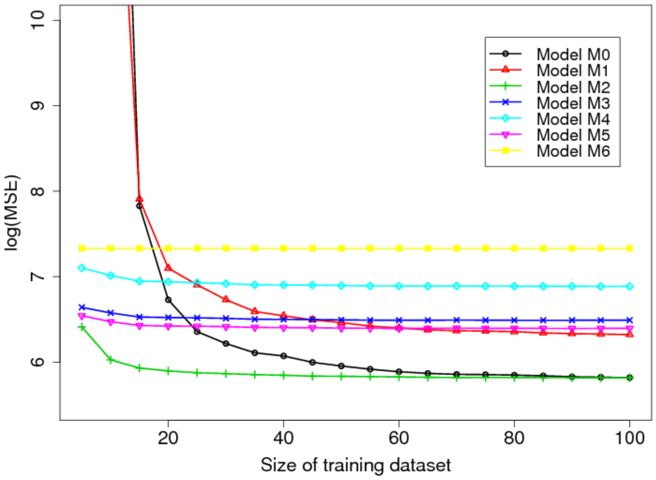

Figure 3. MSE results for the Birmingham-San Jose data.

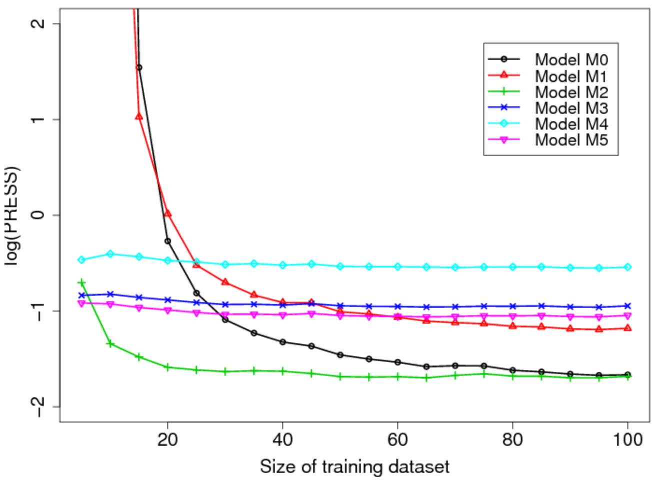

Figure 4. PRESS criterion for the Birmingham-San Jose data.

the San Jose samples, is particularly disappointing (large values of MSE) if it is learned with a very small number of observations and becomes more efficient for learning datasets larger than 50 observations. The model M1 has a similar behaviour for small learning datasets but turns out to be less interesting than M0 when the size of the learning dataset is larger. These behaviours are not surprising since both models M0 and M1 are very complex models and then need large datasets to be correctly learned. Conversely, the models M2 to M5 appear not to be sensitive to the size of the dataset used for adapting the Birmingham model.

In particular, model M2 obtains very low MSE values for a learning dataset size as low as 20 observations. This indicates that model M2 is able to adapt the Birmingham model to San Jose with only 20 observations. Moreover Table 3 indicates that model M2 provides better prediction results than model M0 for the San Jose housing market for learning dataset sizes less than 100 observations. Naturally, since model M0 is more complex,

Table 3. MSE results for the Birmingham-San Jose data.

Model	10 obs.	25 obs.	50 obs.
Model M_0	$3.5\text{x}10^7$	576.9	386.1
Model M_2	**414.8**	**356.7**	**342.1**
Model M_6	1528.9	1528.9	1528.9
Model	100 obs.	250 obs.	all obs.
Model M_0	336.8	**310.7**	**297.5**
Model M_2	**336.0**	332.5	330.1
Model M_6	1528.9	1528.9	1528.9

it becomes more efficient than model M2 for larger datasets even though the difference is not so big for large learning datasets. Figure 4 shows that the PRESS criterion, which will be used in practice since it is computed without a validation dataset, allows the practitioner to successfully select the most appropriate transfer model. Indeed, it appears clearly that the PRESS curves are very similar to the MSE curves computed on the whole dataset. Finally, in such a context, the transformation parameters obtained by the different adaptive linear models can be interpreted in an economic way and this could be interesting for economists. In particular, the estimated transformation parameters by the model M2 with the whole San Jose dataset are $\lambda_0 = 1.439$ and $\lambda = 0.447$. The fact that the San Jose's intercept is almost 50% larger than the one of Birmingham suggests that the minimal basis price of an housing is more expensive in San Jose than in Birmingham. However, the fact that the regression coefficients associated to the explanatory variables of San Jose are on average 50% smaller than the one of Birmingham could mean that the growing of the price according to the housing features is more moderated.

To summarize, this experiment has shown that the adaptive linear models are able to transfer the knowledge on the housing market of a reference city to the market of a different city with a small number of observations. Furthermore, the interpretation of the estimated transformation parameters could help the practitioner to analyze in an economic way the differences between the studied populations.

4.2 The CO_2-GNP data

A mixture of second order polynomial regressions seems to be particularly well adapted to fit the link between CO_2 emissions and the log-GNP, and will be used in the sequel. For the 1980 data, two groups of countries are easily distinguishable: a first minority group (group1, about 25% of the whole sample) is made up of countries for which a growth in the GNP is linked to a high growth in CO_2 emissions, whereas the second group (group 2, about 75%) seems to have more environmentally friendly policies. This country discrimination into two groups is more difficult to obtain on the 1999 data: it seems that countries which had high CO_2 emissions in 1980 have adopted a more environmentally friendly development than in the past, and a two-component mixture regression model could be more difficult to exhibit.

In order to address this issue, adaptive mixture models are used to estimate the mixture regression model on the 1999's data. The eight models pMM2a to pMM3d (since pMM4a and pMM4b are equivalent to pMM2a and pMM2c for p=1), a classical mixture of second order polynomial regressions with two components (MR) and a usual second order polynomial regression (UR) are considered. Different sample sizes of the 1999 data are tested: 30%, 50%, 70% and 100% of the S* size (n*=111). The experiments are repeated 20 times in order to average the results. Table 4 summarizes these results: MSE corresponds to the mean square error, whereas PRESS and BIC are the model selection criteria introduced in Section 3. In this application, the total number of available data in the 1999 population is not sufficiently large to separate them into two training and test samples. For this reason, MSE is computed on the whole S* sample, although a part of it has been used for the training (from 30% for the first experiment to 100% for the last one). Consequently, MSE is a significant indicator of predictive ability of the model when 30% and 50% of the whole dataset are used as training set since 70% and 50% of the samples used to compute the MSE remain independent from the training stage. However, MSE is a less significant indicator of predictive ability for the two last experiments and the PRESS should be preferred in these situations as indicator of predictive ability.

Table 4 first reveals that the 1999 data are actually made of two components as in the 1980 data since both PRESS and MSE are better for MR (2 components) than for UR (1 component) for all sizes n* of S*. This first result validates the assumption that both the reference population P and the new population P* have the same number K=2 of components, and consequently the use of adaptive mixture of regression makes sense for this data. Second, adaptive mixture models turns out to provide very satisfying predictions for all values of n* and in particular outperform the other approaches when n* is small. Indeed, BIC, PRESS and MSE all testify that these models provide better predictions than the other studied methods when n* is equal to 30%, 50% and 70% of the whole sample. Furthermore, it should be noticed that adaptive mixture models provide stable results with respect to variations on n*. In particular, the models pMM2 are those which appear to be the most efficient on this dataset and this means that the link between both populations P and P* is mixture component independent. This application is a good illustration of the advantage of

Table 4. MSE on the whole 1999 sample, PRESS and the BIC criterion for the 8 adaptive mixture models (pMM_{2a} to pMM_{3d}), the usual regression model (UR) and the classical regression mixture model (MR), for 4 sizes of the 1999 sample: 33, 55, 77 and 111 (whole sample). Lower BIC, PRESS and MSE values for each sample size are in bold typeface.

30% of the 1999 data (n*=33)			
model	BIC	PRESS	MSE
pMM_{2a}	13.21	**4.01**	4.77
pMM_{2b}	12.89	4.57	**3.66**
pMM_{2c}	**12.57**	4.16	4.55
pMM_{2d}	17.13	4.38	4.77
pMM_{3a}	15.92	4.49	4.66
pMM_{3b}	16.01	5.59	4.11
pMM_{3c}	15.75	6.17	4.23
pMM_{3d}	22.72	4.49	4.66
UR	27.08	7.46	7.66
MR	32.89	5.54	5.11

50% of the 1999 data (n*=55)			
model	BIC	PRESS	MSE
pMM_{2a}	14.10	4.76	3.88
pMM_{2b}	**13.99**	**4.10**	**3.77**
pMM_{2c}	14.07	5.29	4.22
pMM_{2d}	17.82	4.45	4.66
pMM_{3a}	18.07	4.27	4.66
pMM_{3b}	18.00	5.62	4.44
pMM_{3c}	17.60	5.62	4.33
pMM_{3d}	26.61	6.12	4.55
UR	20.87	7.95	7.21
MR	39.69	4.82	4.77

70% of the 1999 data (n*=77)			
model	BIC	PRESS	MSE
pMM_{2a}	15.15	5.51	8.21
pMM_{2b}	14.82	**3.89**	3.77
pMM_{2c}	**14.71**	4.53	4.44
pMM_{2d}	19.00	5.83	4.99
pMM_{3a}	18.96	4.79	4.44
pMM_{3b}	19.06	4.34	4.22
pMM_{3c}	18.98	5.26	3.77
pMM_{3d}	27.57	5.55	4.88
UR	22.08	8.00	7.10
MR	43.91	5.06	**3.33**

(n*=111)			
model	BIC	PRESS	MSE
pMM_{2a}	15.51	3.83	3.77
pMM_{2b}	15.54	3.87	4.77
pMM_{2c}	**15.34**	4.13	4.11
pMM_{2d}	20.14	4.41	4.33
pMM_{3a}	20.19	4.48	4.77
pMM_{3b}	20.03	4.41	4.33
pMM_{3c}	20.06	4.35	3.44
pMM_{3d}	29.55	4.76	5.44
UR	23.62	7.53	6.99
MR	47.19	**3.66**	**2.89**

combining information on both past (1980) and present (1999) situations in order to analyze the link between CO_2 emissions and gross national product for several countries in 1999, especially when the number of observations for the present situation is not sufficiently large. Moreover, the competition between the adaptive mixture models is also informative. It seems that three models are particularly well adapted to model the link between the 1980 data and the 1999 data: pMM2a, pMM2b and pMM2c. The particularity of these models is that they consider the same transformation for both classes of countries, which means that all the countries have the same kind of evolution.

The estimated mixture of two regression models on the 1980 data is:

$$CO_2 = 26.96\text{-}9.62\log(GNP) + 0.88\log(GNP)2$$
$$CO_2 = 13.42\text{-}4.57\log(GNP) + 0.40\log(GNP)2$$

with respective probabilities $\pi_1 = 0.26$ and $\pi_2 = 0.74$ and residual variances $\sigma_1^2 = 3.10$ and $\sigma_2^2 = 0.55$. The model

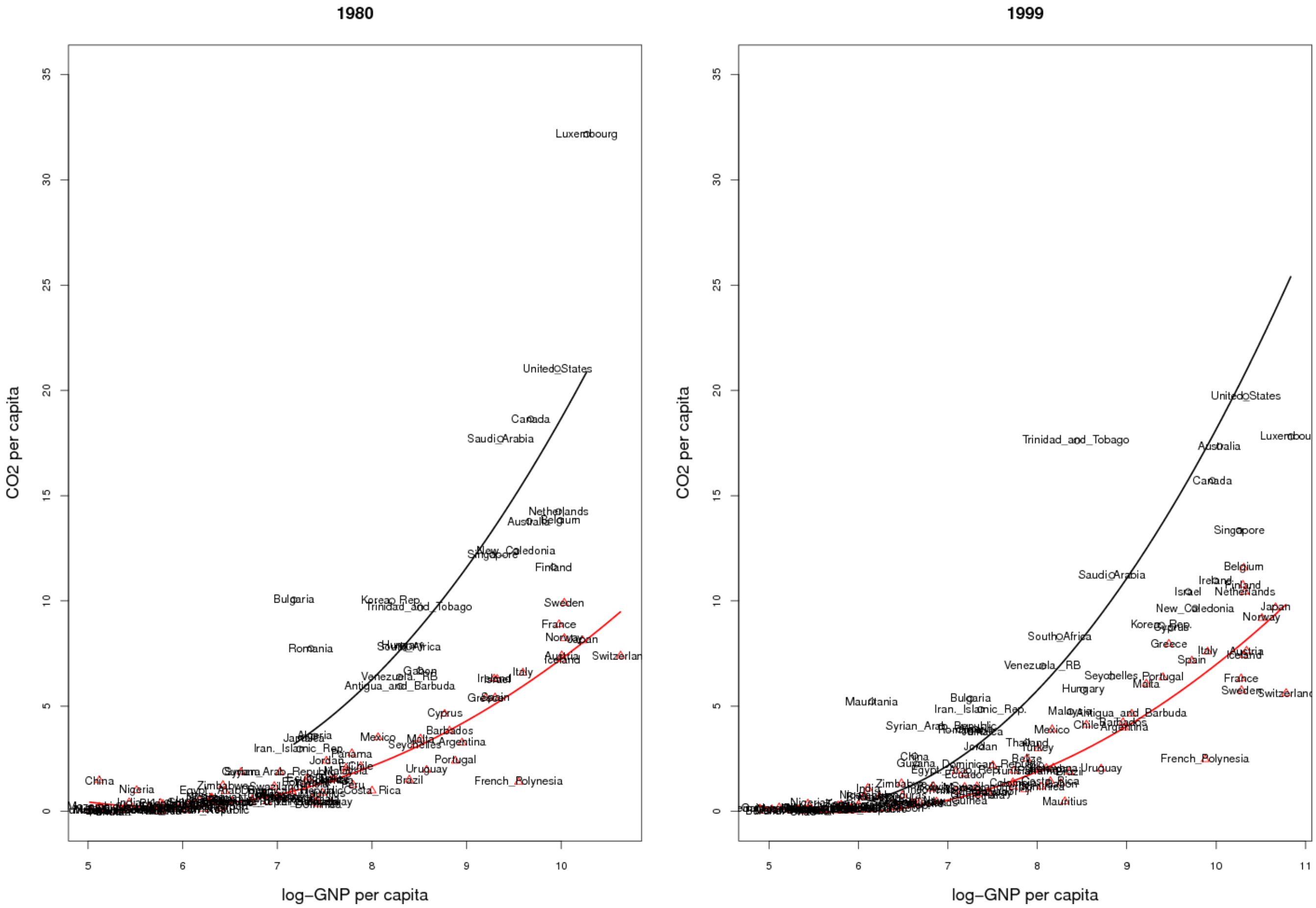

Figure 5. Emissions of CO_2 per capita versus GNP per capita in 1980 (left) and 1999 (right) and estimated adaptive mixture models (with model pMM_{2c} for 1999).

for the 1999 data estimated with model pMM2c (for the whole sample size) is obtained with a link parameter $\lambda = 1.26$:

$$CO_2 = 33.92 - 12.1\log(GNP) + 1.11\log(GNP)2$$
$$CO_2 = 16.89 - 5.75\log(GNP) + 0.50\log(GNP)2$$

with $\pi_1 = 0.15$, $\sigma_1^{*2} = 4.9$, $\pi_2 = 0.85$ and $\sigma_2^{*2} = 0.87$. These results are illustrated in Figure 5.

5. Discussion

When carrying out a regression analysis to analyze a phenomenon which has already been studied but in different conditions, adaptive models can help exploit the previous analysis in order to emphasize the quality of the current one. In this paper, we have shown how a regression model predicting the house values can be adapted from the US South to the US West Coast, and how the regression of CO_2 emissions in terms of gross national product in 1999 can be estimated by using information about the same analysis in 1980.

Similar models exist also in the context of classification tasks (Biernacki et al. 2002, Jacques and Biernacki 2010).

REFERENCES

Allen, D.M. 1974. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics,* 16:125-127.

Barker, T. 1991. Measuring economic costs of CO_2 emission limits. *Energy,* 16(3):611-614.

Biernacki, C., Benimel, F. & Bretagnolle, V. 2002. A generalized discriminant rule when training population and test population differ on their descriptive parameters. *Biometrics,* 58(2):338-397.

Bouveyron, C. & Jacques, J. 2010. Adaptive linear models for regression: improving prediction when population has changed. *Pattern Recognition Letters*, 31(14):2237-2247. doi:10.1016/j.physletb.2003.10.071.

Bouveyron, C. & Jacques, J. 2010. Adaptive mixtures of regressions: Improving predictive inference when population has changed. *Pub. IRMA Lille*, 70(8):1-25.

Dempster, A.P., Laird, N.M. & Rubin, D.B. 1977. Maximum likelihood from incomplete data (with discussion). *Journal of the Royal Statistical Society. Series B*, 39:1-38.

Goldfeld, M. & Quandt, R.E. 1973. A Markov model for switching regressions. *Journal of Econometrics*, 1:3-16.

Grubb, M., et al. Ha-Duong, M. 1997. Influence of socio-economic inertia and uncertainty on optimal CO_2-emission abatement. *Nature*, 390:170-173.

Hurn, M., Justel, A. & Robert, C.P. 2003. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, 12(1):55-79.

Jacques, J. & Biernacki, C. 2010. Extension of model-based classification for binary data when training and test populations differ. *Journal of Applied Statistics*, 37(5):749-766.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461-464.

Correspondence: charles.bouveyron@univ-paris1.fr