Quantitative Economics for the Evaluation of the European Policy

Dipartimento di Economia e Management

Co-funded by the Erasmus+ Programme of the European Union

Project funded by

European Commission Erasmus + Programme –Jean Monnet Action

Project number 553280-EPP-1-2015-1-IT-EPPJMO-MODULE

Irene Brunetti Davide Fiaschi Angela Parenti[1]

26/10/2017

[1] ireneb@ec.unipi.it, davide.fiaschi@unipi.it, and aparenti@ec.unipi.it.

Econometric model of convergence

Hypothesis of absolute convergence with linear model: $\beta < 0$

$$\overline{g_{Y/L}} = intercept + \beta \log\left(Y/L_{i,1991}\right) + \epsilon_i \tag{1}$$

	Dependent variable:
	c(averageGrowthRateSingleRegions)
LogGDPpwRel_initialYear	-0.014^{***}
	(0.001)
Constant	0.015^{***}
	(0.001)
Observations	257
R^2	0.506
Adjusted R^2	0.504
Residual Std. Error	0.010 (df = 255)
F Statistic	261.512^{***} (df = 1; 255)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Some Remarks on Generalize Additive Models (GAMs)

A generalize additive model is a generalized linear model with a linear predictor involving a smooth function of the covariate:

$$y_i = s(x_i) + \epsilon_i \tag{2}$$

where y_i is the response variable, x_i the covariate, $s(\cdot)$ a smooth function and ϵ_i are i.i.d. $N(0, \sigma^2)$.

Some Remarks on Generalize Additive Models (GAMs)

A generalize additive model is a generalized linear model with a linear predictor involving a smooth function of the covariate:

$$y_i = s(x_i) + \epsilon_i \tag{2}$$

where y_i is the response variable, x_i the covariate, $s(\cdot)$ a smooth function and ϵ_i are i.i.d. $N(0, \sigma^2)$.
To estimate $s(\cdot)$, let represent it is such a way that Eq. (2) becomes a linear model by choosing a *basis* function:

$$s(x) = \sum_{j=1}^{q} b_j(x)\beta_j \tag{3}$$

for some values of the unknown parameters.

Some Remarks on Generalize Additive Models (GAMs)

A generalize additive model is a generalized linear model with a linear predictor involving a smooth function of the covariate:

$$y_i = s(x_i) + \epsilon_i \tag{2}$$

where y_i is the response variable, x_i the covariate, $s(\cdot)$ a smooth function and ϵ_i are i.i.d. $N(0, \sigma^2)$.
To estimate $s(\cdot)$, let represent it is such a way that Eq. (2) becomes a linear model by choosing a *basis* function:

$$s(x) = \sum_{j=1}^{q} b_j(x)\beta_j \tag{3}$$

for some values of the unknown parameters.

For example, consider a 4th order polynomial where
$b_1(x) = 1, b_2(x) = x, b_3(x) = x^2, b_4(x) = x^3, b_5(x) = x^4$, so that Eq.(3) becomes:

$$s(x) = \beta_1 + x\beta_2 + x^2\beta_3 + x^3\beta_4 + x^4\beta_5. \tag{4}$$

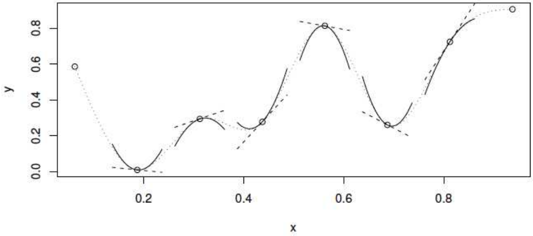
Some Remarks on Generalize Additive Models: Cont.

Polynomial bases have some problems in estimating on the whole domain of x. Alternatively, the **spline** bases perform well.

Some Remarks on Generalize Additive Models: Cont.

Polynomial bases have some problems in estimating on the whole domain of x. Alternatively, the **spline** bases perform well.

A cubic spline is a curve, made up of sections of cubic polynomial, joined together so that they are continuous in value as well as first and second derivatives. The points at which the sections join are known as the **knots** of the spline



Knots must be chosen!

Some Remarks on Generalize Additive Models: Cont.

Given knot location, how do we choose the **degree of smoothing** (i.e. the basis dimension)?

One way, is to keep the basis dimension fixed, at a size a little larger than it is believed could reasonably be necessary, but to control the models smoothness by adding a wiggliness penalty to the least squares fitting objective.

Some Remarks on Generalize Additive Models: Cont.

Given knot location, how do we choose the **degree of smoothing** (i.e. the basis dimension)?

One way, is to keep the basis dimension fixed, at a size a little larger than it is believed could reasonably be necessary, but to control the models smoothness by adding a wiggliness penalty to the least squares fitting objective.

For example, rather than fitting the model by minimizing:

$$\| \mathbf{y} - \mathbf{X}\beta \|^2 \tag{5}$$

it could be fit by minimizing,

$$\| \mathbf{y} - \mathbf{X}\beta \|^2 + \lambda \int [s''(x)] dx \tag{6}$$

where the integrated square of second derivative penalizes models that are too wiggly.

Some Remarks on Generalize Additive Models: Cont.

Given knot location, how do we choose the **degree of smoothing** (i.e. the basis dimension)?

One way, is to keep the basis dimension fixed, at a size a little larger than it is believed could reasonably be necessary, but to control the models smoothness by adding a wiggliness penalty to the least squares fitting objective.

For example, rather than fitting the model by minimizing:

$$\| \mathbf{y} - \mathbf{X}\beta \|^2 \tag{5}$$

it could be fit by minimizing,

$$\| \mathbf{y} - \mathbf{X}\beta \|^2 + \lambda \int [s''(x)] dx \tag{6}$$

where the integrated square of second derivative penalizes models that are too wiggly. The trade off between model fit and model smoothness is controlled by the *smoothing parameter*, λ. $\lambda \to \infty$ leads to a straight line estimate for s, while $\lambda = 0$ results in an un-penalized regression spline estimate.

Some Remarks on Generalize Additive Models: Cont.

Given that s is linear in parameters β the penalty can be written as a quadratic form:

$$\lambda \int [s''(x)]dx = \beta' \mathbf{S} \beta \tag{7}$$

where $\mathbf{S}$ is a matrix of known coefficients.

Therefore the penalized regression spline problem is to minimize:

$$\| \mathbf{y} - \mathbf{X}\beta \|^2 + \lambda \beta' \mathbf{S} \beta \tag{8}$$

$$\rightarrow \hat{\beta} = (\mathbf{X}'\mathbf{X} + \lambda \mathbf{S}')^{-1} \mathbf{X}'\mathbf{y} \tag{9}$$

λ must be chosen!! Too high λ leads to over-smoothing while too low λ to under-smoothing $\rightarrow$ generalized cross-validation.

Some Remarks on Generalize Additive Models: Cont.

Both with the over-smoothing (too high λ) and the under-smoothing (too low λ) the spline estimate $\hat{f}$ will not be close to the true function f.

Ideally, it would be good to choose λ so that $\hat{f}$ is as close as possible to f.

A suitable criterion might be to choose λ to minimize:

$$M = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_i - f_i)^2 \tag{10}$$

Problem: f is no observable!

Some Remarks on Generalize Additive Models: Cont.

It is possible to derive an estimate of $E(M) + \sigma^2$ which is the expected squared error in predicting a new variable.

Let $\hat{f}^{-i}$ be the model fitted to all data except y_i and define the *ordinary cross validation* score:

$$V_0 = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}^{-i} - y_i \right)^2 \tag{11}$$

This score results from leaving one out each datum in turn, fitting the model to the remaining data and calculating the squared difference between the missing datum and its predicted value: these suared differences are the averaged over all data

Some Remarks on Generalize Additive Models: Cont.

It is computationally inefficient to calculate V_0 by leaving one out one datum at a time, and fitting the model to each of the n resulting datasets.

But it can be shown that:

$$V_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}_i)^2/(1 - A_{ii})^2 \tag{12}$$

where $\hat{f}$ is the estimate from fitting all the data and A is the corresponding influence matrix.

In practice, the weights $(1 - A_{ii})$ are often replaced by the mean weights $tr(I - A)/n$, in order to arrive at the *generalized cross validation*.

Some Remarks on Generalize Additive Models: Cont.

When we have and additive model with more than one covariate as for example:

$$y_i = s_1(x_i) + s_2(z_i) + \epsilon_i \tag{13}$$

the parameters β are obtained by minimization of the penalized least squares objective:

$$\| \mathbf{y} - \mathbf{X}\beta \|^2 + \lambda_1 \beta' \mathbf{S_1} \beta + \lambda_2 \beta' \mathbf{S_2} \beta \tag{14}$$

Econometric model of convergence (cont.d)

Hypothesis of absolute convergence with a nonparametric model: $s' < 0$

$$\overline{g_{Y/L}} = intercept + s\left(\log\left(Y/L_{i,1991}\right)\right) + \epsilon_i \tag{15}$$

Parametric coeff.:	Estimate	Std. Error	t-Stat	P-value
(Intercept)	0.0175179	0.0006103	28.7	$< 2e^{-16}$ ***
Smooth terms:	edf	Ref.df	F	p-value
s(LogGDPpwRel_initialYear)	8.722	8.978	37.97	$< 2e^{-16}$ ***
	R-sq.(adj)=0.565; Dev.expl.=58% GCV=$9.948e^{-05}$; Scale est.=$9.5717e^{-05}$; n=257			

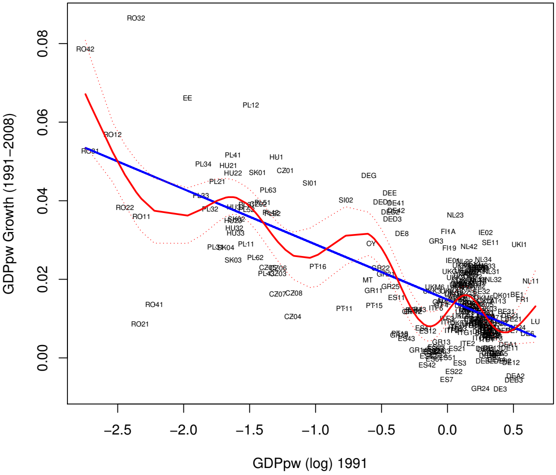


Figura: Absolute convergence in the GDP per worker. Parametric and nonparametric regression

Conditional convergence

Hypothesis of conditional convergence with linear model: $\beta_0 < 0$

$$\overline{g_{Y/L}} = intercept + \beta_0 \log\left(Y/L_{i,1991}\right) + \beta_1 \bar{s} + \beta_2 \bar{n} + \beta_3 \bar{h} + \epsilon_i \tag{16}$$

	Dependent variable:
	c(averageGrowthRateSingleRegions)
LogGDPpwRel_initialYear	-0.015^{***}
	(0.001)
log(mean_invRate)	0.003
	(0.003)
log(mean_empGR_augm)	-0.015^{***}
	(0.003)
log(mean_hc_index)	0.020^{***}
	(0.002)
Constant	-0.093^{***}
	(0.012)
Observations	257
R^2	0.641
Adjusted R^2	0.635
Residual Std. Error	0.009 (df = 252)
F Statistic	112.557^{***} (df = 4; 252)
Note:	*p<0.1; **p<0.05; ***p<0.01

Conditional convergence

Hypothesis of conditional convergence with a nonparametric model: $s_0' < 0$

$$\overline{g_{Y/L}} = intercept + s_0\left(\log\left(Y/L_{i,1991}\right)\right) + s_1\left(\bar{s}\right) + s_2\left(\bar{n}\right) + s_3\left(\bar{h}\right) + \epsilon_i \quad (17)$$

Parametric coeff.:	Estimate	Std. Error	t-Stat	P-value
(Intercept)	0.0175179	0.0004611	37.99	$< 2e^{-16}$ ***
Smooth terms:	edf	Ref.df	F	p-value
s(LogGDPpwRel_initialYear)	8.641	8.963	39.175	$< 2e^{-16}$ ***
s(log(mean_invRate))	5.392	6.582	1.722	0.109
s(log(mean_empGR_augm))	8.595	8.95	5.644	$< 2e^{-16}$ ***
s(log(mean_hc_index))	1.235	1.434	80	$< 2e^{-16}$ ***
	R-sq.(adj)=0.752; Dev.expl.=77.5%			
	GCV=$6.0497e^{-05}$; Scale est.=$5.4645e^{-05}$; n=257			

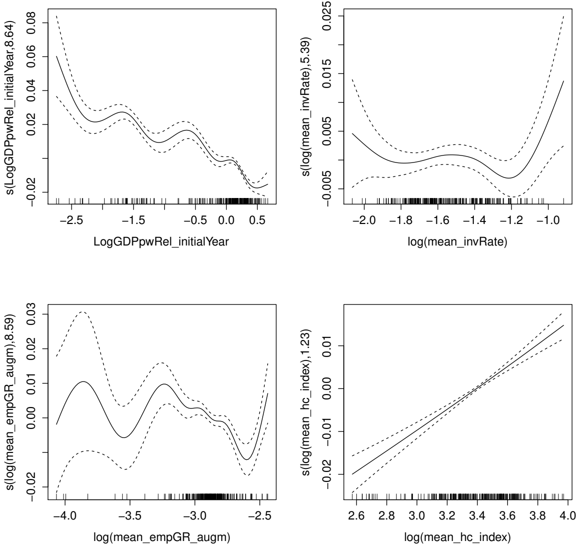


Figura: Estimate of generalized additive model.