Project 2: High-dimensional Linear Models and Convergence in Economic Growth

AØKA08084U Advanced Microeconometrics

University of Copenhagen

Character Count (incl. spaces): 11962

Sebastian Hørlück, Matias Piqueras

November 7, 2021

1 Introduction

This assignment analyses if countries that start out as poor tend to have a faster growing GDP compared to richer countries when controlling for confounding. In the literature, this is referred to as conditional β-convergence and we can formalize an econometric model to approach the problem as

$$g_i = \beta y_{i0} + \boldsymbol{z}_i' \boldsymbol{\gamma} + u_i \tag{1}$$

where g_i is the average annual growth rate of GDP for country i, y_{i0} the logarithm of initial GDP per capita, and $\boldsymbol{z}_i$ is a p-dimensional vector of control variables where potentially $p >> n$. In the following of this paper, we will let uppercase symbols denote matrices of variables or parameters such that e.g. G denotes the matrix of g_i for all observations $i \in 1, 2, ..., n$.

The rest of the text is structured as follows: first we detail the econometric theory allowing for consistent estimation of β and its asymptotic variance. Next, we discuss and show how an initial variable selection is performed. In this respect, we are guided by both the econometric literature on economic growth and estimation in a high-dimensional setting. Finally, the results of our study are presented before discussing the limitations and weaknesses of our approach.

2 Methods for estimating β in a high-dimensional framework

When estimating a linear relationship, such as the one presented in (1), a apparent approach could be to use the OLS-estimator. In lower-dimensional frameworks, where $\frac{p}{n} \to 0$, as $n \to \infty$, the OLS estimator performs well under relatively mild assumptions. In our case, we cannot let $n \to \infty$, as we have only a limited number of countries, and further we want to allow for a number of variables relatively high compared to n, such that $\frac{p}{n} \to \infty$ is nonneglible. As shown in (Riis-Vestergaard Sørensen, 2021a) the expected estimation error of $\hat{\beta}^{OLS}$ is:

$$\begin{aligned} \mathrm{E}\left[\sum_{j=1}^{p}\left(\hat{\beta}_j^{OLS} - \beta_j\right)^2\right] &= \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{p}\left[X_j'\left(\hat{\beta}_j^{OLS} - \beta\right)\right]\right] \\ &= \sigma^2\frac{p}{n} \end{aligned}$$

thus it performs badly when $\frac{p}{n}$ is nonneglible. Thus we diverge from the idea of using OLS to estimate β from (1), and instead introduce a model which allows us to include a large set of control variables, where we can let $p >> n$.

2.1 Sparsity

Choosing a subset of the variables for estimating β is legitimate if we believe the relation between G and $\mathbf{Z}$ is sparse. That is that only a few of the parameters related to the variables in $\mathbf{Z}$ are non-zero such that

$$s = ||\gamma||_0 = \sum_{j=1}^{p} \mathbf{1}\{\gamma_j \neq 0\} \text{ is small}$$

This exact sparsity condition can be formulated in a less restrictive form, as the approximate sparsity condition. This condition states that a lot of γ's can be non-zero as long as only a few γ's are far from zero while the rest are close to zero.

2.2 LASSO-estimation

Using the sparsity assumption stated in section 2.1, we introduce the LASSO model.

$$\hat{\beta}(\lambda) := \underset{b \in \mathbf{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(g_i - \beta y_{i0} - \mathbf{z_i}'\gamma \right)^2 + ||b||_1 \right\} \tag{2}$$

where $||b||_1 := |\beta| + \sum_{j=1}^{p} |\gamma_j|$. The LASSO-methods effectively shrinks estimates until the point of extinction, effectively dropping variables that do not add substantial information in the form of explained variance. Thus, this approach relies on the exact sparsity condition, as outlined in section 2.1.

It is worth noting, that the dropping of variables can lead to strong omitted variable bias, if we are in an approximate sparse situation, compared to the exact sparsity condition the LASSO requires. This is seen through the example of variables strongly correlated with y_{i0} but not adding substantial information to the prediction of g_i being dropped through the LASSO-method. To circumvent these challenges, we need to select an appropriate level of λ, which we describe further in section 2.3.

There does not exist an analytical representation of $\hat{\beta}$, thus we cannot draw any conclusions regarding the inference of $\hat{\beta}$. To estimate confidence intervals, we introduce the Post Double-LASSO (PDL).

2.2.1 Post Double-LASSO

In short, PDL selects the relevant controls in $\mathbf{z_i}$, by using LASSO on a estimation of both (1) and y_{i0} using $\mathbf{z_i}$, that is we introduce a second regression model:

$$y_{i0} = \mathbf{z_i}'\psi + \nu_i \tag{3}$$

Combining the residuals from (1) and (3) respectively, investigate the following moment:

$$G(\beta, \eta) = (y_{i0} - \mathbf{z_i}'\psi)(g_i - \beta y_{i0} - \mathbf{z_i}'\gamma), \eta = \{\psi, \gamma\} \tag{4}$$

It can be shown that for a given $\beta = \beta_0$, the moment above is orthogonal to η, see (Riis-Vestergaard Sørensen, 2021c). Thereby we construct the following moment condition:

$$\begin{gathered} \mathrm{E}\left[(y_{i0} - \mathbf{z_i}'\psi)(g_i - \beta y_{i0} - \mathbf{z_i}'\gamma)\right] = 0 \Leftrightarrow \\ \frac{\mathrm{E}\left[(y_{i0} - \mathbf{z_i}'\psi)(g_i - \mathbf{z_i}'\gamma)\right]}{\mathrm{E}\left[(y_{i0} - \mathbf{z_i}'\psi)y_{i0}\right]} = \beta \end{gathered} \tag{5}$$

Which leads to the PDL β-estimator:

$$\frac{\sum_{i=1}^{N}\left[(y_{i0} - \mathbf{z_i}'\hat{\psi})(g_i - \mathbf{z_i}'\hat{\gamma})\right]}{\sum_{i=1}^{N}\left[(y_{i0} - \mathbf{z_i}'\hat{\psi})y_{i0}\right]} = \hat{\beta}^{PDL} \tag{6}$$

Following Belloni, Chernozhukov, and Hansen (2014b), it proves that $\sqrt{(}N)(\hat{\beta} - \beta)$ converges in distribution to a normal distribution with mean zero and variance, σ^2. Further, (6) suggests the estimated variance of $\hat{\beta}^{PDL}$ as:

$$\hat{\sigma}^2 = \frac{N^{-1}\sum_{i=1}^{N}\hat{u}_i^2\hat{\nu}_i^2}{\left(N^{-1}\sum_{i=1}^{N}\hat{\nu}_i^2\right)^2}$$

Thus PDL suggests an appropriate method for estimating β and constructing the related confidence intervals.

2.2.2 Post Partialling Out-LASSO

Building on the reasoning for (4), and introducing a regression of GDP growth on the controls alone:

$$g_i = \mathbf{z_i}'\delta + \xi_i \tag{7}$$

we construct the following moment from (7) and (3):

$$G(\beta, \tilde{\eta}) = (y_{i0} - \mathbf{z_i}'\psi)((g_i - \mathbf{z_i}'\delta) - \beta\,(y_{i0} - \mathbf{z_i}'\psi))\,, \tilde{\eta} = \{\psi, \delta\} \tag{8}$$

Just like with the PDL-estimator, it can be shown that (8) is orthogonal to $\tilde{\eta}$, motivating the following moment condition:

$$\frac{\mathrm{E}\left[(y_{i0} - \mathbf{z_i}'\psi)(g_i - \mathbf{z_i}'\delta)\right]}{\mathrm{E}\left[(y_{i0} - \mathbf{z_i}'\psi)^2\right]} = \beta \tag{9}$$

leading us to the PPOL β-estimator:

$$\frac{\sum_{i=1}^{N}\left[(y_{i0}-\mathbf{z_i}'\hat{\psi})(g_i-\mathbf{z_i}'\hat{\gamma})\right]}{\sum_{i=1}^{N}\left[(y_{i0}-\mathbf{z_i}'\hat{\psi})^2\right]}=\hat{\beta}^{PPOL} \tag{10}$$

Likewise as with the PDL-estimator, an appropriate confidence interval for $\hat{\beta}^{PPOL}$ can be constructed by estimating the variance

$$\hat{\sigma}^2=\frac{N^{-1}\sum_{i=1}^{N}\hat{\xi}_i^2\hat{\nu}_i^2}{\left(N^{-1}\sum_{i=1}^{N}\hat{\nu}_i^2\right)^2}$$

thus allowing for the construction of the related confidence intervals (Riis-Vestergaard Sørensen, 2021c).

2.3 Choosing λ

λ controls the strength of the regularization and is a hyperparameter that needs to specified. In the statistical learning literature, this is often done by choosing the λ that minimizes the prediction error on a validation set using for example 5-fold cross-validation (Friedman, 2017). While this approach is reasonable if the goal is prediction, it lacks theoretical justification and is therefore likely to be a bad choice when we are interested in learning something about the models parameters (Chetverikov & Sørensen, 2021). For this reason we will rely on the work by Belloni, Chen, Chernozhukov, and Hansen (2012) and what they refer to as the feasible procedure (hereafter we will refer to it as BCCH) for picking λ. They show that it yields asymptotically well behaved estimates for the IV based estimators introduced in section 2.2.1 under the assumption of sparsity even when p is much larger than n. Further, they show that λ^{BCCH} also dominate the score $\lambda^{BCCH}\geq c||S||_\infty$ with probability at least $1-\alpha$ (see also Riis-Vestergaard Sørensen (2021b)), such that the estimation errors are bounded. The score is defined as:

$$S:=\frac{2}{n}\sum_{i=1}^{n}[u_iX_i]$$

To find the λ^{BCCH} we first need to chose the parameters $\alpha\in(0,1)$ which we set to $\alpha=0.05$ and c which we set to the typical value $c=1.1$ following Riis-Vestergaard Sørensen (2021a). Next, we use the data to estimate a pilot penalty level $\widehat{\lambda}^{\text{pilot}}$ as

$$\widehat{\lambda}^{\text{pilot}} := \frac{2c}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2p}\right)\max_{1\leqslant j\leqslant p}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2 X_{ij}^2} \tag{11}$$

where Φ is the standard normal distribution. We then run a pilot Lasso using $\widehat{\lambda}^{\text{pilot}}$ in order to get the residuals $\widehat{\epsilon} := Y_i - X_i'\widehat{\beta}^{\text{pilot}}$. These residuals are then plugged into the last term of equation (12) to yield our BCCH penalty level

$$\widehat{\lambda}^{\text{BCCH}} := \frac{2c}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2p}\right)\max_{1\leqslant j\leqslant p}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 X_{ij}^2}. \tag{12}$$

We want to note two things about the BCCH. First, there exists other methods for picking λ that involve fewer steps than BCCH, like for example the Bickel-Ritov-Tsybakov Rule. However, it relies on σ of the residuals ϵ being known and that ϵ is homoscedastic. σ could be conservatively approximated by $\sqrt{\frac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{N}}$ but risks being imprecise. More importantly, we are unwilling to assume homoscedasticity. Doing so would imply for example that the variance of GDP growth is the same for the smallest and largest countries. As we discuss further in section 3, we want to allow for large amount of controls of potentially unknown functional form to enter the model. Imposing that all of these should be homoscedastic is therefore likely to be an unreasonable assumption and unnecessary given the option to use BCCH. Secondly, equations (11) and (12) are written generally for pairs (Y_i, X_i), however, for both PDL and PPOL we must estimate $\widehat{\lambda}^{\text{BCCH}}$ separately for the pairs $(g_i, \boldsymbol{z}_i)$ and $(y_i, \boldsymbol{z}_i)$.

3 Variable Selection

Whilst the list of determinants of economic growth is potential very long the number of countries that we can study is finite and relatively small. To name a few candidate controls, the quality of a country's institutions may play a considerable role in explaining income differences. Acemoglu, Johnson, and Robinson (2001), Acemoglu, Naidu, Restrepo, and Robinson (2019), and Ashraf and Galor (2013) all seek explanations of low GDP among countries, focusing on institutions, democratization, and genetic diversity respectively. Moreover, we might think that other factors such as a countries' access to natural resources, religiosity and propensity to experience natural disasters all, to some extent, affect economic growth. When faced with the task of isolating the effect of initital

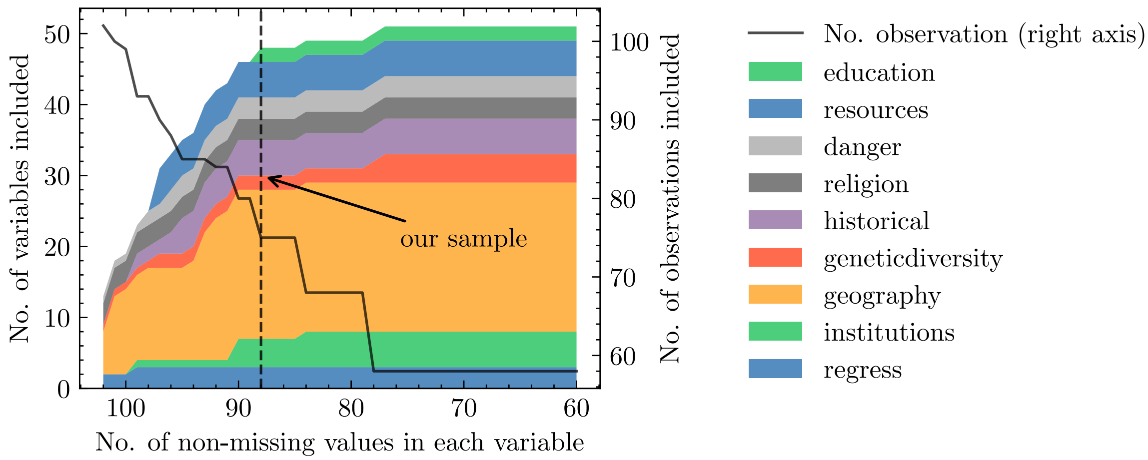

Figure 1: Trade-off between variables and observations

GDP per capita, we therefore need to make some kind of selection of what variables to include in the model. In Belloni, Chernozhukov, and Hansen (2014a) they refer to this as regularization and note that this can either be done manually by the researcher based on economic intuition or by letting the data itself guide the selection of variables. The first approach suffers from many researcher degrees of freedom as it so heavily relies on domain knowledge to avoid omitted variable bias and moreover knowing "True" the functional form of the underlying model, if for example interaction terms polynomials should be included. Problems of misspecification by omitting relevant controls has explicitly been pointed in relation the the literature on economic growth (Sala-i Martin, 1997).

Against this background and given our limited domain knowledge, we deem it necessary to allow for a maximum number of variables while keeping the amount of observations at a reasonable level. In figure 1 we plot this trade-off and as can seen be number of observations without missing values decreases as a function of the number of non-missing values in each variable. We think that 75 observations and 48 variables (including GDP growth and initial GDP) strikes a good balance as it allows us to include variables from each category and still recover a fair amount of observations. Table 3 in the Appendix lists the included variables.

Prioritizing more observations would lead us to drop the education category of control variables, which we deem important to control for as cognitive skills on the population level has been shown to be strongly related to economic growth (Hanushek & Wößmann, 2007). Figure 2 further shows the countries that are included in our analysis and Table 2 in the appendix the countries we have dropped relative to the sample of 102 countries. As can

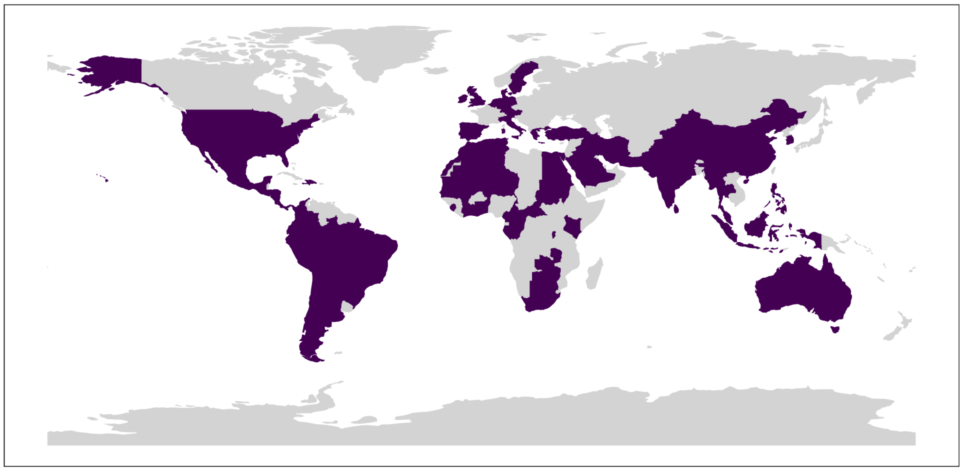

Figure 2: Included Countries

be seen both Africa (mostly west and central) and the Americas are over-represented. To avoid strong assumptions of functional form, we include interactions between all controls and second-order polynomials, resulting in a total of 1128 variables.

4 Results

We show the results from PDL in Table 1. In line with the conditional β-convergence hypothesis, our estimate for the standardize logarithm of initial GDP shows that one standard deviation increase leads to a -0.166% decrease in GDP growth per capita. Back-of-the-envelope calculations suggests that countries with one negative standard deviation in inital GDP will have an approximate 8.3% increase in GDP per capita over a 50 year period relative to a country at the mean.

However, since the upper bound of the 95% confidence interval is positive we are unable to reject the null-hypothesis of no β-convergence. In other words, we cannot based on these results, say that countries who start out relatively poorer exhibit a stronger GDP growth. We only report the results for PDL since PPOL led to the exact same result up to the fifth decimal point. Similar results is to be expected since the methods are second-order equivalent (Riis-Vestergaard Sørensen, 2021c).

Table 1: Post-Double-Lasso Results

	Dependent variable:
	Percentage GDP growth
Log initial GDP	-0.166
	(-0.430; 0.099)
Observations	75
Number of controls	46 (1,128)
Interactions and squares	YES
Number of non-zero control coefficients	0

Note: 95% confidence intervals in parenthesis. λ is calculated according to the BCCH-rule. Number of controls incl. second-order polynomials and interactions in parenthesis. Number of non-zero controls is based on the coefficients from the Lasso $g_i = \boldsymbol{z}_i'\gamma + u_i$ that are non-zero.

5 Discussion

Though our results show some indications of β-convergence, we are not able to reject the null hypothesis of no effect from initial GDP on GDP growth. There are several points which challenge our results.

Firstly, in our aim to include as many control variables as possible, we are also faced with the choice of omitting more observations, since the data quality aross different countries vary a lot. Even though our LASSO approach allows for the estimation of β with more variables than observations, the low number of observations also causes a relatively low statistical power. In our regression analysis we include 75 observations, which opens up the possibility of including more countries, thus improving the statistical power of the analysis, with improved data coverage.

Secondly, related to the argument above, the resulting set of countries in our analysis might not be representative of countries in general. Following Figure 2, we see that we might lack representation of African and east-European countries. The level of representation could be investigated further by looking at the distribution of variables in countries that are not included relatively to included countries. A possible finding of lacking representativeness could be mitigeted by weighing included countries with low representativeness relative to countries with high representativeness.

References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, *91*(5), 1369–1401.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy does cause growth. *Journal of political economy*, *127*(1), 47–100.

Ashraf, Q., & Galor, O. (2013). The'out of africa'hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, *103*(1), 1–46.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, *80*(6), 2369–2429.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014a). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, *28*(2), 29–50.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014b). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, *81*(2), 608–650.

Chetverikov, D., & Sørensen, J. R.-V. (2021). Analytic and bootstrap-after-cross-validation methods for selecting penalty parameters of high-dimensional m-estimators. *arXiv preprint arXiv:2104.04716*.

Friedman, J. H. (2017). *The elements of statistical learning: Data mining, inference, and prediction.* springer open.

Hanushek, E. A., & Wößmann, L. (2007). The role of education quality for economic growth. *World Bank policy research working paper*(4122).

Riis-Vestergaard Sørensen, J. (2021a). *Linear model in high dimensions, 1: Introduction (annotated).* University Lecture.

Riis-Vestergaard Sørensen, J. (2021b). *Linear model in high dimensions, 2: Implementation and theory (annotated).* University Lecture.

Riis-Vestergaard Sørensen, J. (2021c). *Linear model in high dimensions, 3: Inference (annotated).* University Lecture.

Sala-i Martin, X. (1997). *I just ran four million regressions.* National Bureau of Economic Research Cambridge, Mass., USA.

Appendix

Table 2: Missing Countries

Code	Country	Continent
BFA	Burkina Faso	Africa
TCD	Chad	Africa
SYC	Seychelles	Africa
SWZ	Eswatini	Africa
NGA	Nigeria	Africa
MWI	Malawi	Africa
GNB	Guinea-Bissau	Africa
MDG	Madagascar	Africa
HKG	Hong Kong	Asia
BGD	Bangladesh	Asia
GEO	Georgia	Asia
SGP	Singapore	Asia
PNG	Papua New Guinea	Australia
KIR	Kiribati	Australia
FJI	Fiji	Australia
LUX	Luxembourg	Europe
MLT	Malta	Europe
FIN	Finland	Europe
TTO	Trinidad and Tobago	North America
BHS	Bahamas	North America
BLZ	Belize	North America
VCT	Saint Vincent and the Grenadines	North America
PRI	Puerto Rico	North America
JAM	Jamaica	North America
URY	Uruguay	South America
SUR	Suriname	South America
GUY	Guyana	South America

Table 3: Included Variables

category	**variable**	**label**
regress	gdp_growth	Annual growth in GDP per capita, 1970-2020
regress	lgdp_initial	GDP per capita in 1970 (log)
regress	investment_rate	Capital formation (% of GDP per year, avg. of ...
institutions	dem	Democracy measure by ANRR
institutions	demCGV	Democracy measure by CGV
institutions	demBMR	Democracy measure by BMR
institutions	demreg	Average democracy in the region*initial regime...
geography	tropicar	% land area in geographical tropics
geography	distr	mean distance to river
geography	distcr	mean distance to coast or river
geography	distcr	Mean distance to nearest waterway
geography	distc	mean distance to coast
geography	suitavg	Land suitability for agriculture
geography	temp	Temperature
geography	suitgini	Land suitability Gini
geography	elevavg	Mean elevation
geography	elevstd	Standard deviation of elevation
geography	kgatr	Percentage of population living in tropical zones
geography	precip	Precipitation
geography	area	Total land area
geography	abslat	Absolute latitude
geography	cenlong	Geodesic centroid longitude
geography	area_ar	Arable land area
geography	rough	Terrain roughness
geography	landlock	=1 if landlocked
geography	africa	Africa dummy
geography	asia	Asia dummy
geography	oceania	Oceania dummy
geography	americas	Americas dummy
geneticdiversity	pdiv	Predicted genetic diversity
geneticdiversity	pdiv_aa	Predicted genetic diversity (ancestry adjusted)
historical	pd1000	Population density in 1000 CE

historical	pd1500	Population density in 1500 CE
historical	pop1000	Population in 1000 CE
historical	pop1500	Population in 1500 CE
historical	ln_yst	Log [Neolithic transition timing]
religion	pprotest	Share of Protestants in the population
religion	pcatholic	Share of Roman Catholics in the population
religion	pmuslim	Share of Muslims in the population
danger	yellow	=1 if vector yellow fever present today
danger	malfal	Percentage of population at risk of contractin...
danger	uvdamage	Ultraviolet exposure
resources	oilres	oil reserves
resources	goldm	Natural minerals: gold
resources	iron	Natural mineral: iron
resources	silv	Natural mineral: silver
resources	zinc	Natural mineral: zinc
education	ls_bl	Percentage of population with at most secondar...
education	lh_bl	Percentage of population with tertiary educati...