Advanced Microeconometrics Project 2 1
0pt

University of Copenhagen

Advanced Microeconometrics - Project 2

Signe Skovgaard Klysner, Asbjørn Juul Petersen og Emma Houmøller Veng

07.11.2021

1 Introduction

Conventional long term growth models such as the Solow and Ramsay model imply conditional convergence hypothesizing that initially poor countries tend to "catch up" to richer by having a relatively high growth rate. This corresponds to obtaining a negative point estimate when regressing growth on initial GDP and controls also referred to as β-convergence. This has fueled a long standing debate on what causes growth. To name a few examples, Acemoglu, Johnson, and Robinson (2001) argue that institutions and democracy are the main drivers, Ashraf and Galor (2013) put emphasis on the importance of genetic diversity. With many candidate controls within the literature and relative few countries we investigate the theory of β-convergence in a high dimensional setting. We conduct a post double LASSO estimation with a penalty level determined by the Belloni-Chen-Chernozhukov-Hansen (BCCH) rule and Cross validation (CV) to determine which variables we should control for. We find negative point estimates of initial GDP on growth implying a "catching-up" effect of initially poor countries. The effect is only significant using CV.

2 Data

We utilize a sample of 214 countries which contains information about growth rates, initial level of GDP and a long list of candidate control variables. We perform an initial variable selection among the candidate controls, as even high dimensional methods can break down with too many variables included. We include candidate controls which contain information about geography, education, and natural resources. Additionally we include a variable for population growth, investment rate and a democracy measure. The variables are chosen as they are frequently mentioned as possible drivers for economic growth (e.g. Diamond, 1997).

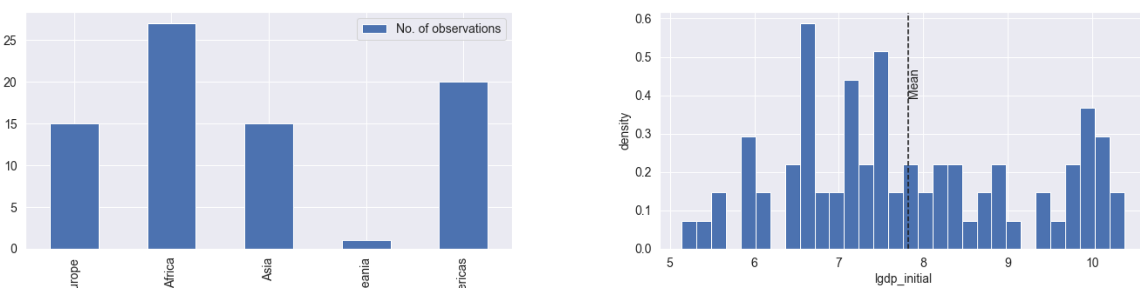
We exclude countries with missing values. Based on our initial variable selection our sample contains 78 countries and 31 candidate controls. We add interaction terms across the candidate controls to increase the flexibility of the model. We exclude all interactions with no variation. This provides us with a high dimensional sample with 482 candidate controls. For a full list of the included variables see the appendix.

Figure 1(a) depicts the geographical distribution of the countries included in the sample. This shows that African countries are over-represented in relation to Asian countries.

Figure 1: Summary statistics of selected sample

(a) Continents

(b) Initial LogGDP



3 Econometric Theory

In order to test the convergence hypothesis we wish to estimate the following econometric model of interest:

$$g_i = \beta_o y_{i0} + \boldsymbol{z_i}\boldsymbol{\gamma} + u_i = \boldsymbol{X'}\boldsymbol{\Phi} + u_i \tag{1}$$

g_i is the growth rate in GDP per capita. y_{i0} is the log GDP per capita in 1970. $\boldsymbol{z_i}$ is a vector of candidate controls, which is included to ensure that the model does not suffer from omitted variable bias. The β-convergence hypothesis amounts to testing whether $\beta < 0$, i.e. that initially poor countries catch up to initially richer countries. $\boldsymbol{X_i}$ and $\boldsymbol{\Phi}$ are vectors of combined regressors and coefficients respectively used to simplify notation.

The issue is to select which candidate controls to include in $\boldsymbol{z_i}$. If we include all $p = 483$ regressors OLS breaks down as the rank condition is not fulfilled. Even if we include less regressors than the amount of countries n but $\frac{p}{n}$ is large then OLS performs poorly as it can be shown that the prediction error is equal to:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{X_i'}\hat{\mathbf{\Phi}}^{\mathbf{LS}} - \mathbf{X_i'}\mathbf{\Phi})^2\right] = \frac{\sigma^2 p}{n} \tag{2}$$

3.1 The LASSO Estimator

The LASSO Estimator performs regularization and variable selection and is therefore useful in a high dimensional framework as ours. A key condition for LASSO is the assumption of sparcity. I.e. we assume that only a few of the coefficients are different from zero:

$$s = ||\mathbf{\Phi}||_0 = \sum_{j=1}^{p} \mathbf{1}\{\phi_j \neq 0\} \text{ is small} \tag{3}$$

Under this assumption the LASSO estimator is defined as:

$$\hat{\Phi}(\lambda) = \underset{b \in R^p}{argmin} \left\{\frac{1}{n}\sum_{i=1}^{n}(g_i - \mathbf{X_i}'b)^2 + \lambda||b||_1\right\}, \quad ||b||_1 = \sum_{j=1}^{p}|b_j| \tag{4}$$

b is the vector of candidate coefficients. The idea is to shrink the least square estimator by adding a cost of complexity λ to the minimization problem. Intuitively the LASSO estimator adds a penalty for the absolute value of every non-zero coefficient.

3.2 Selection of penalty parameter, λ

To select λ we apply both a prediction driven approach and a rule-driven approach. The group of prediction driven approaches include sample splitting and CV. They select λ to minimize out of sample prediction errors. This usually result in more included variables.

The rule driven approaches include the Bickel-Ritov-Tsybakov (BRT) rule and the Belloni-Chen-Chernozhukov-Hansen (BCCH) rule. They are more focused on variable selection, which generally result in a higher λ. We use one from each class of approaches to investigate whether our results are robust across different procedures for choosing λ. We select CV over sample splitting as sample splitting does not make optimal use of the data. We select BCCH over BRT as the BRT rule unrealistically requires homoskedasticity and known variance of the error term.

3.2.1 Cross-validation

The procedure of cross-validation splits the sample into K subsamples of equal size. For each subsample $k = 1, 2..., K$ we use subsample k for validation and the others for training. We then calculate the squared out-of-sample prediction error for each subsample k. Cross-validation then chooses λ by minimizing the sum of misfits:

$$\hat{\lambda}^{CV} = \underset{\lambda}{\operatorname{argmin}} \sum_{k=1}^{K} F_k(\lambda) \tag{5}$$

Where $F_k(\lambda)$ is the average squared prediction error using subsample k as validation sample.

3.2.2 Belloni-Chen-Chernozhukov-Hansen Rule (BCCH)

To apply the BCCH rule we first select a parameter α which roughly can be interpreted as the confidence level of our estimator. We choose $\alpha = 0.05$. Secondly we choose a mark-up c. We use $c = 1.1$. Next we run a pilot-Lasso with the penalty level:

$$\hat{\lambda}^{pilot} = \frac{2c}{\sqrt{n}} \Phi^{-1} \left(1 - \frac{\alpha}{2p}\right) \sqrt{\underset{1 \leq j \leq p}{max} \frac{1}{n} \sum_{i=1}^{n} (g_i)^2 \mathbf{X_{ij}}^2} \tag{6}$$

After that we calculate the residuals:

$$\hat{\epsilon}_i = g_i - \mathbf{X_i}' \hat{\mathbf{\Phi}}^{pilot}, \tag{7}$$

and run a final Lasso with the following λ:

$$\hat{\lambda}^{BCCH} = \frac{2c}{\sqrt{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2p}\right)\sqrt{\underset{1\leq j\leq p}{max}\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i{}^2\mathbf{X_{ij}}^2} \tag{8}$$

3.3 Post-Double Lasso & Inference

We estimate the parameter associated with initial growth by Post-Double Lasso (PDL). The Lasso estimator in equation 4 is biased towards zero. Further, the asymptotic distribution is unknown. The most intuitive approach would be to use LASSO-estimation to select among controls and subsequently estimate equation (1) by OLS with the selected controls. However, This approach suffers from omitted variable bias as Lasso may exclude non-zero variables with moderate effects (Benolli et. al 2014).

The idea behind PDL is that we impose some additional structure to our problem. We consider the augmented model:

$$g = \beta_o y_0 + \boldsymbol{z\gamma_o} + u \;\text{ where } E[u|y_0, \boldsymbol{z}] = 0 \tag{9}$$

and

$$y_0 = \boldsymbol{z\psi_o} + \upsilon \;\text{ where } E[\upsilon|\boldsymbol{z}] = 0 \tag{10}$$

$\boldsymbol{\gamma_0}, \boldsymbol{\phi_0}$ are nuisance parameters. I.e. parameters of no interest but which we are required to estimate to obtain an estimate of β_o. Equation (10) is an added "first stage" that regress the candidate controls on the initial GDP. Equations (9) and 10 imply the following moment condition:

$$E[(y_0 - \boldsymbol{z\psi_o})(g - \beta_o y_0 - \boldsymbol{z\gamma_o})] = 0 \tag{11}$$

It is shown in (slide 27) that small errors in the nuisance parameters have little to no impact on the parameter of interest and the moment condition is orthogonalized wrt. the nusisance parameters. Thus we can identify the true parameter of interest as:

$$\beta_o = \frac{E[(y_0 - \boldsymbol{z\psi_o})(g - \boldsymbol{z\gamma_o})]}{E[(y_0 - \boldsymbol{z\psi_o})y_0]} \tag{12}$$

We estimate this by first using LASSO to estimate g_i on y_{i0} and $\boldsymbol{z_i}$ (long-regression). This helps us obtain $\hat{\beta}$ and $\hat{\boldsymbol{\gamma}}$. Next we use LASSO to estimate y_{i0} on $\boldsymbol{z_i}$ (first-stage). This helps us obtain $\hat{\boldsymbol{\psi}}$. We use this to consistently estimate β_o using the analogy principle by replacing expectations and the true parameters from equation 12 with averages and parameter estimates:

$$\hat{\beta}^{PDL} = \frac{\sum_{i=1}^{n}(y_{i0} - \boldsymbol{z_i}\hat{\boldsymbol{\psi}})(g_i - \boldsymbol{z_i}\hat{\boldsymbol{\gamma}})}{\sum_{i=1}^{n}(y_{i0} - \boldsymbol{z_i}\hat{\boldsymbol{\psi}})y_{i0}} \tag{13}$$

Under certain conditions the Post-Double Lasso is consistent and asymptotically normal as $n \to \infty$, even in high dimensions, with distribution:

$$\frac{\sqrt{n}(\hat{\beta}^{PDL} - \beta_o)}{\sigma_o} \xrightarrow{d} N(0,1), \quad \sigma_o^2 = \frac{E[\epsilon^2 \upsilon^2]}{(E[\upsilon^2])^2} \tag{14}$$

We can estimate σ_o^2 per the analogy principle:

$$\hat{\sigma}^2 = \frac{n^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i^2\hat{\upsilon}_i^2}{(n^{-1}\sum_{i=1}^{n}\hat{\upsilon}_i^2)^2} \tag{15}$$

In order to test our hypothesis of β-convergence we construct a confidence interval based on $\hat{\beta}^{PDL}$ and $\hat{\sigma}^2$:

$$\hat{CI}_{1-\xi} = \left[\hat{\beta}^{PDL} \pm q_{1-\xi/2}\frac{\hat{\sigma}}{\sqrt{n}}\right] \tag{16}$$

Where q is the standard normal quantile function. ξ is the chosen level of significance. It can be shown that:

$$P(\beta_o \in \hat{CI}_{1-\xi}) \to 1 - \xi \tag{17}$$

In this paper we choose $\xi = 0.05$. The theory for β-convergence amounts to testing whether the confidence interval only contains negative parameter values.

4 Results

We estimate model 1 with Post-Double Lasso outlined in section 3.3. We estimate the model both by using CV and BCCH to select the penalty parameter λ. The results are reported in table 1

Table 1: Post Double Lasso Results

	CV	BCCH
$\hat{\beta}^{PDL}$	-0.705	-0.154
se	0.351	0.133
λ_{Xg}	0.128	2.578
λ_{Zy_0}	0.047	1.230
Confidence interval, lower bound	-1.394	-0.414
Confidence interval, upper bound	-0.017	0.106
Number of observations	78	78
Number of variables	483	483

λ_{Xg} is the penalty level used when we Lasso g on X

λ_{Zy_0} is the penalty level used when we Lasso y_0 on Z

The results show that $\hat{\beta}^{PDL}$ is negative both when we select the penalty parameter by CV and by BCCH. The magnitude of the estimates, however, are different for the two approaches. The estimated effects using CV implies that a one standard deviation increase in the log of initial GDP decreases the subsequent growth with around 0.7 standard deviations. This effect is statistically significant. When we use BCCH the estimated effect is around -0.15 standard deviation and not statistically different from zero.

These differences stem from the amount of selected control variables in the two cases: CV has selected lower penalty parameters. As a consequence 19 non-zero variables are selected in the "long-regression" and 33 non-zero variables are selected in the "first-stage". Intuitively this high number of variables are selected as CV keeps adding variables as long as it increases the prediction power.

The BCCH algorithm has selected much higher penalty parameters both for the "long regression" and the "first stage". As a result no variables (except from the intercept) have non-zero coefficients in either of the regressions. It is apparent from equation 13 that in this special case $\hat{\beta}^{PDL}$ simplifies to the estimate of a simple linear regression of g_i on y_{i0}.

To sum up, the selected amount of controls with CV and BCCH represent two extremes. In reality neither of the cases are probably realistic. We find a negative point estimate of β in both cases. The effect is only significantly different from zero using CV. These results partially support the convergence hypothesis.

5 Discussion & Concluding Remarks

A possible concern is that our model suffers from omitted variables bias. This is the case if the conditional independence assumptions stated in equation (9) and (10) are violated. This is most likely the case when the BCCH-rule applies. It seems unrealistic that no candidate controls are correlated with initial GDP or growth. One example of such a violation is the resource curse (kilde?): This theory implies that countries with many natural resources initially have a high level of GDP but fail to build solid institutions which result in a lower subsequent growth. As the BCCH has selected no variables natural resources is in the error term u_{it}. This theory invalidates the moment condition in equation (11) and thus our result.

CV selects more variables, however, this provides no guarantees against omitted variable bias. If our initial variable selection is insufficient CV could still result in omitted variables bias. We have not included any measures of e.g. genetic diversity in the list of candidate controls. It is likely correlated with both initial GDP and subsequent growth.

Another possible source of endogeneity is sample selection. A valid concern is that we only have data for countries who are rich now meaning that we could potentially systematically have under sampled countries with a low initial GDP and a low subsequent growth. This would bias our results towards acceptance of the growth hypothesis. However, the fact that Africa is in fact over sampled in this analysis somewhat alleviates this concern.

Another concern is that LASSO performs poorly when the regressors are highly correlated. This might the case in our setting as we have allowed for many interaction terms. One way to mitigate this issue would be to use the Elastic Net estimator, which is a combination between Ridge Regression and Lasso. The benefit of this estimator is that it both performs variable selection and provides stability when regressors are highly correlated.

6 Literature

Daron Acemoglu, Simon Johnson, and James A. Robinson. The colonial origins of comparative development: An empirical investigation. *American Economic Review,* 91(5):1369-1401, 2001.

Jared Diamond. Guns, Germs, and Steel. W. W. Norton (1st edition). 1997

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

INFERENCE ON TREATMENT EFFECTS AFTER SELECTION AMONGST HIGH-DIMENSIONAL CONTROLS

7 Appendix

Variables included in the sample:

'tropicar', 'distr', 'distcr', 'distc', 'suitavg', 'temp', 'suitgini', 'elevavg', 'elevstd', 'kgatr', 'precip', 'area', 'abslat', 'cenlong', 'area_ar', 'rough', 'landlock', 'africa','asia', 'oceania', 'americas', 'oilres', 'goldm', 'iron', 'silv', 'zinc', 'ls_bl', 'lh_bl', 'pop_growth', 'investment_rate', 'dem',