Assignment 2

Mathias P., Mikkel R., Noah C.

Faculty of Social Sciences

University of Copenhagen

1 Introduction

Macroeconomic theory suggests that if we regress the average growth rate of GDP per capita g_i in the period from 1970 to 2020 on logged GDP per capita $y_{i,0}$ in 1970, we would observe a negative relationship between the two. Theoretically, this is often attributed to countries with low initial GDP having a large growth potential by a.o. accumulating physical and human capital, while countries with high initial GDP has already been through this process. This concept is formally denoted β-convergence

The regression equation of interest is (1).

$$g_i = \beta y_{i,0} + \boldsymbol{z}_i \boldsymbol{\gamma} + u_i \tag{1}$$

where $\boldsymbol{z}_i$ is a $p-1$ vector of control variables and u_i is the unobservable error term with $\mathbb{E}[u_i | y_{i,0}, \boldsymbol{z}_i] = 0$.

We follow the reasoning in Barro (1989), that stresses the importance of including relevant controls to avoid endogenity issues. Barro (1989) observes very low correlation between g_i and $y_{i,0}$ when not including proxys for e.g. human capital. With this in mind, we consider including a plethora of available control variables in $\boldsymbol{z}_i$

Our dataset contains observations of g_i and $y_{i,0}$ for (only) $n = 102$ countries and including control variables implies that n might decrease, while p increases. This issue is hard to avoid since we don't know the relevant control variables of $\boldsymbol{z}_i$ a-priori. This leaves us with the task of control variable selection in a High-Dimensional (HD) setting, while retaining the possibility of inference. We formally define the HD setting as $\frac{p}{n} \not\to 0$, namely that we deal with datasets where the number of regressors p is not incomparable to the number of observations n. *In our setting, we end up with* $p = 604 \gg 88$, *implying that e.g. OLS is infeasible.*

In the following we consider data-driven control variable selection using the Least Absolute Shrinkage and Selection Operator (LASSO) and discuss inference in that refard. This largely resolves the issues presented in Sala-I-Martin (1997), namely that running multiple regressions might emphasize importance of a variable in one regression and drop the same variable in a different regression.

The null hypothesis we seek to test is:

$$H_0 : \hat{\beta} < 0, \quad \text{vs} \quad H_A : \hat{\beta} \geq 0$$

2 Methodology

While OLS is feasible when $p < n$, it is not suitable for HD variable selection as it overfits. The OLS estimator minimizes the squared residuals of n observations wrt. p regressors. In the case of $n = p$, OLS perfectly (over-)fits the n observations with $p = n$ degress of freedom. This leads us to consider LASSO. Like OLS, LASSO minimizes squared residuals and a penalty term in the form of the ℓ_1-norm of the coefficients. For (1), the LASSO is:

$$\begin{bmatrix}\hat{\beta}(\lambda)\\ \hat{\boldsymbol{\gamma}}(\lambda)\end{bmatrix} \in \arg\min_{\beta\in\mathbb{R},\gamma\in\mathbb{R}^{p-1}} \frac{1}{n}\sum_{i=1}^{n}(g_i - \beta y_{i,0} - \boldsymbol{z}_i\boldsymbol{\gamma})^2 + \lambda\left(|\beta| + \|\hat{\boldsymbol{\gamma}}\|_1\right) \tag{2}$$

Using LASSO in itself for estimation of β in (1) is not meaningful for inference as the asymptotic distribution of $\hat{\beta}(\lambda)$ is unknown. Further, when LASSO chooses the coefficient for a variable, it balances the explanatory power of the variable with model complexity, which biases coefficients toward zero.

Instead, we consider Post-Single-LASSO (PSL), Post-Double-LASSO (PDL), Post-Partialling-Out-LASSO (PPOL) and Post-Double-Selection-LASSO (PDS), which all use LASSO as a variable selection device and then adhere to OLS or moment conditions for inferential estimation.

2.1 Consistency of LASSO and penalty selection

The first task is choosing an appropriate penalty λ, which relates to the consistency of LASSO. We can't derive the distributional properties of LASSO, but it is possible to derive a deterministic error bound $\|\hat{\beta}(\lambda) - \beta\|_2$ as a function of λ.

Assumption of approximate sparsity (AS): While the model is HD, we assume that it is approximately sparse $\frac{s}{n} \to 0$, where s denotes the number of non-zero coefficients picked by LASSO. Formally, this means that $\mathbb{E}[g_i|\boldsymbol{z}_i]$ is well approximated by $s \ll n$ important control variables.

We consider two rules for choosing λ. The first one is the Bickel-Ritov-Tsybakov (BRT) rule from Bickel et al. (2009) with $\hat{\lambda}^{BRT}$, which relies on $u_i|y_{i,0}, \boldsymbol{z}_i \sim \mathcal{N}(0, \sigma^2)$. This leads us to consider the robust Belloni-Chen-Chernozhukov-Hansen (BCCH) rule choosing $\hat{\lambda}^{BCCH}$, which is described in detail in Belloni et al. (2012). Under the BRT and BCCH rule, the error bound can be shown to be vanishing with probability 1 ($\alpha_n \to 0$) if $\frac{s\ln(p)}{n} \to 0$ when $n \to \infty$. Both penalties are increasing in p: $\lambda^{BRT} = 2c\sigma\Phi^{-1}(1 - a/2p)$, and similarly for λ^{BCCH}, and therefore LASSO might select fewer regressors when we expanding the amount of regressors via. polynomials.

A third option for choosing λ is the Cross-Validation (CV) scheme, which is preferred for prediction but selects to many variables. This leads us to proceed with the BCCH-rule but also implement CV and BRT for comparison and highlighting differences. For a detailed description of consistency in the case of CV, we refer to Chetverikov et al. (2020).

2.2 Estimation based on LASSO

One way of using LASSO as a variable selection device is to use PSL, which estimates (2), gather the non-zero coefficient variables in $\boldsymbol{z}_i$ and then estimates (1) using only these variables by OLS. However, Belloni et al. (2014) shows that the resulting OLS estimates are not normally distributed due to PSL occasionally dropping relevant control variables as they correlate with $y_{i,0}$. With this in mind, we consider PDL, PPOL and PDS, which are based on two steps. In the first stage of PDL, we LASSO $y_{i,0}$ on $\boldsymbol{z}_i$, obtaining $\hat{\psi}$.

In the second stage of PDL, we LASSO g_i on $y_{i,0}$ and $\boldsymbol{z}_i$. Using the moment condition $\mathbb{E}[u_i \epsilon_i] = 0$, identifies β and leads to the following estimate $\check{\beta}$ using the analogy principle:

$$\check{\beta} = \frac{\sum_{i=1}^{n} (y_{i,0} - \boldsymbol{z}_i' \hat{\psi})(g_i - \boldsymbol{z}_i' \hat{\gamma})}{\sum_{i=1}^{n} (y_{i,0} - \boldsymbol{z}_i' \hat{\psi}) y_{i,0}},$$

which is asymptotically normal. For PPOL we LASSO g_i on only $\boldsymbol{z}_i$ and not $y_{i,0}$, thereby obtaining a different estimate for $\hat{\gamma}$. While PPOL directly estimates β based on LASSO residuals, it is at risk of excluding control variables that predict both g_i and $y_{i,0}$. However, PDL and PPOL share features as e.g. normality of $\check{\beta}$ and are asymptotically equivalent. PDS estimates $\check{\beta}$ using OLS on $y_{i,0}$ and the union of controls selected from the first and second selection steps. The PDS estimator is also asymptotically normal and the interested reader is referred to Belloni et al. (2014) for a thorough description.

3 Empirical analysis

3.1 Variable selection and data description

The data consists of 85 variables for 214 countries of which 102 are used. We use the pre-curated list provided with 46 additional candidate regressors.

Using all 46 variables without any interpolation would limit the sample size to $n = 63$. This leads us to include 42 of the 46[1], limiting n to 88, only interpolating the institutional dummy variables related to democracy[2]. This classifies the problem as one of HD. We justify our choices in accordance with classical growth theory and results from Barro (1989) and Sala-I-Martin (1997) who stresses the importance of including a.o. regional, political and religious variables.. Following the framework of classical growth models as Solow, Ramsey-Cass-Koopmans and extensions of the two, we include population growth and the investment rate as baseline controls reflecting capital accumulation and demographic

[1] List of candidate regressors in appendix.

[2] For only 5 countries where we assume a missing value is equal to the country being undemocratic

dynamics. Beyond these, we include groups of variables capturing geography, institutions, religion, and historical factors. Lastly, since we are approximating a potentially highly unlinear function $\mathbb{E}[g_i|\boldsymbol{z}_i]$, we include polynomial features[3]. This brings the number of candidate regressors up to 604 such that $p \gg n$, in which LASSO is essential in selecting relevant regressors under the assumption of sparsity.

3.2 Results

Table 1 presents baseline OLS estimates. The first estimate is not significant, while the second is significantly negative. Including polynomial features makes $p \gg n$ in which OLS is infeasible[4], and furthermore, polynomial expansion introduces severe rank deficiency. As such, the OLS estimates provides inconsicive conclusions with respect to β-convergence at best. This follows the notion from Barro (1989), namely that estimating $\hat{\beta}$ without relevant controls show small correlation between g_i and $y_{i,0}$.

Figure 1 shows LASSO-paths from which we see the trade-off between penalization methods in variable selection as outlined in 2 in line with the assumption of sparsity. Similarly PSL consistently chooses fewer regressors, but we cannot perform inference due to $\hat{\beta}$ being non-normally distributed. and BRT is deemed unreliable due to non-robustness.

The selected variables differ across methods, as well as candidate regressor set. Without polynomial controls, disregarding CV, BRT selects `asia`, `uvdamage`, `demreg` (Stage 1), `asia`, `malfal`, `pop1500` (Stage 2) as explanatory variables wrt. g_i, for both PDL, PPOL, and PDS. BCCH selects none for PSL, and only `uvdamage` for the rest. Including polynomial features changes the selected variables, with BRT selecting $\texttt{abslat} \cdot \texttt{demreg}$, $\texttt{uvdamage} \cdot \texttt{pdiv}$ (Stage 1), and $\texttt{cenlong} \cdot \texttt{pop1500}$, $\texttt{malfal}^2$, `asia` (Stage 2), in all models[5]. Most puzzling is that with the expanded candidate set, BCCH selects zero variables, which we discuss in Section 4.

All procedures (except for CV PSL) used in Table 2 provide negative point estimates. PDL and PPOL provide significant estimates using CV, and insignificant for BRT and BCCH, while PDS provides consistently negative estimates for all penalizations methods. Due to BCCH being the most robust penalization method for our purposes, we find mixed results with respect to β-convergence with non-polynomial set.

Lastly, Table 2 provides estimates using a candidate regressor set with polynomial features. While estimates are largely consistent[6] with those of Table 1, now with a clear

[3]For all non-binary variables we include squared terms, interaction terms, as well as squared interaction terms

[4]We have used the pseudo-inverse to remedy the numerical problems.

[5]The latter is also the selected variables from PSL.

[6]Estimates are within |0.2| for BRT and BCCH (except BRT PSL) of each other which adds robustness to our conclusions.

rejection of the null of β-convergence. Both BRT and BCCH select fewer variables in which the latter chooses zero. As such, all $\hat{\beta}$ estimates, fall back to that of OLS, simply a regression of $y_{i,0}$ on g_i.

Overall, we do not find compelling evidence that support the hypothesis of β-convergence, that is, poorer countries do not "catch-up" to richer.

4 Discussion and conclusion

In contrast to former econometric analysis of β-convergence as Barro (1989) and Sala-I-Martin (1997), we employ data-driven variable selection procedures to avoid ad-hoc decisions about which variables to include. This contrast does not ensure the LASSO methods selects the correct control variables, and we still had to make choices about which variables to include in the first place. A general econometric concern is that of model selection. Under the assumption of linearity in parameters, this then becomes a question on which variables to include such that the model is the true functional form. PSL assumes that the selected variables, represent true model selection, which is highly unlikely. This leads us to consider PDL, PPOL and PDS instead. When considering the baseline controls versus the polynomial controls for these LASSO methods, we observe LASSO not selecting the same candidate regressors because 1) It is likely that the correlated polynomials distort the signal for LASSO to pick up on the true regressors. 2) Higher penalties ($\lambda^{BCCH} \approx 2.4615$ with polynomials and $\lambda^{BCCH} \approx 1.6971$ without) result in LASSO not selecting the same candidate regressors.

This also leads us to discuss whether the AS assumption is fulfilled. The error bounds for LASSO is based on $\frac{s\ln(p)}{n} \to 0$, implying that LASSO would be inconsistent if $\mathbb{E}[g_i|\boldsymbol{z}_i]$ is not approximated well by few control variables. As economic growth is the result of many different processes, it is hard to completely rule out the case when the AS assumption could be violated. An argument for this could e.g. be that Sala-I-Martin finds 22 variables significant, when examining the same question, using a very different variable selection scheme.

With all this in mind, we find that in the non-polynomial case under BCCH, all estimation procedures besides PSL selects uvdamage as the only relevant control variable. We suspect that this might be due to uvdamage being an "indirect" proxy for multiple elements relevant for growth. uvdamage intrinsically carries a lot of geographical information, possibly being correlated with human-capital.

Overall we do not find evidence that support the hypothesis of β-convergence, that is, a negative relationship between $y_{i,0}$ and g_i, contrary to Barro (1989) .

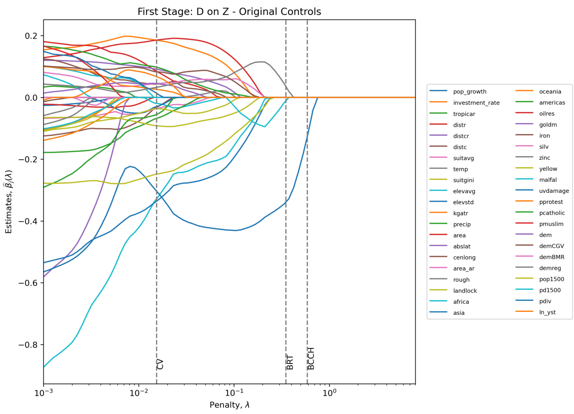
A Appendix

The list of candidate regressors:

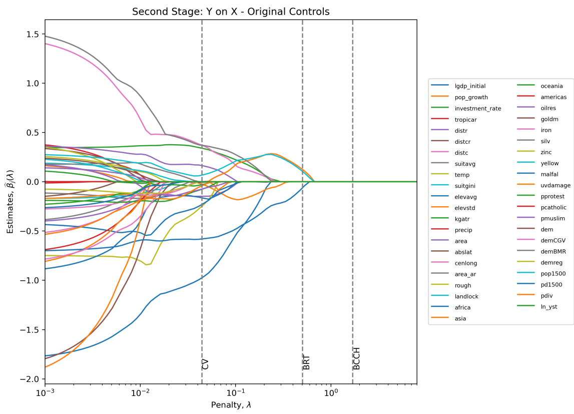
$z = \{$`pop_growth`, `investment_rate`, `tropicar`, `distr`, `distcr`, `distc`, `suitavg`, `temp`, `suitgini`, `elevavg`, `elevstd`, `kgatr`, `precip`, `area`, `abslat`, `cenlong`, `area_ar`, `rough`, `landlock`, `africa`, `asia`, `oceania`, `americas`, `oilres`, `goldm`, `iron`, `silv`, `zinc`, `yellow`, `malfal`, `uvdamage`, `pprotest`, `pcatholic`, `pmuslim`, `dem`, `demCGV`, `demBMR`, `demreg`, `pop1500`, `pd1500`, `pdiv`, `ln_yst`$\}$.

Lasso-paths:

Figure 1: First and second stage estimates using original control specification

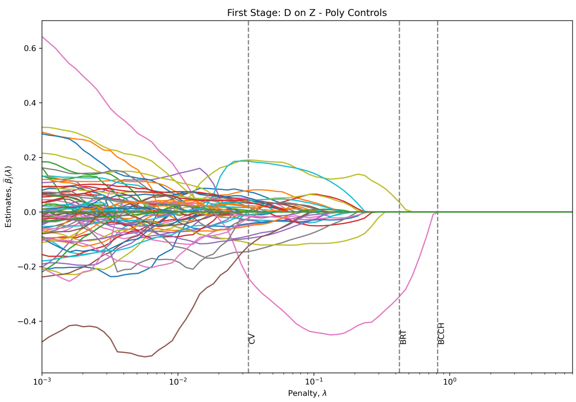


(a) First stage (original controls)

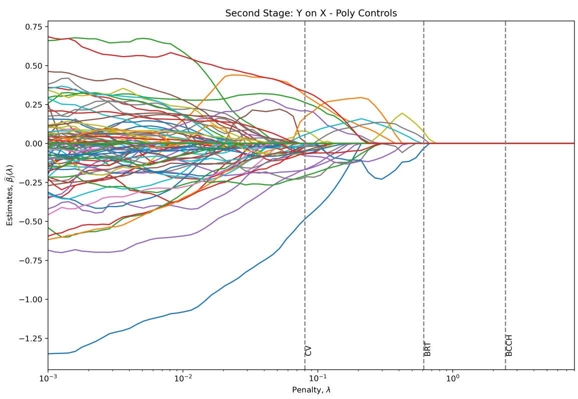


(b) Second stage (original controls)

Figure 2: First and second stage estimates using polynomial feature expansion



(a) First stage (polynomial features)



(b) Second stage (polynomial features)

Estimates

Table 1: OLS Estimation Results

Specification	$\hat{\beta}$	**CI**
No Controls	-0.121	$(-0.5984,\ 0.3566)$
	(0.244)	
With Controls	-1.810^{***}	$(-2.227,\ -1.394)$
	(0.213)	
With Poly	-354.767^{***}	$(-355.222,\ -354.312)$
	(0.232)	

Note: Each model considers $N = 88$. Standard errors appear in parentheses below point estimates. The *With Poly* specification suffers from severe rank deficiency; estimates are computed using a pseudo-inverse and are included for illustrative purposes only. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 2: LASSO Estimation Results (Baseline Controls)

	CV			**BRT**			**BCCH**		
	$\hat{\beta}$	CI	$\dot{\boldsymbol{z}}$	$\hat{\beta}$	CI	$\dot{\boldsymbol{z}}$	$\hat{\beta}$	CI	$\dot{\boldsymbol{z}}$
Post-Lasso Method									
PSL	0.033	$(-0.126,\ 0.192)$	17	-0.324^{**}	$(-0.539,\ -0.109)$	3	-0.121	$(-0.419,\ 0.177)$	0
	(0.081)			(0.110)			(0.152)		
PDL	-1.287^{***}	$(-1.969,\ -0.604)$	42	-0.326	$(-0.669,\ 0.016)$	6	-0.185	$(-0.467,\ 0.096)$	1
	(0.348)			(0.175)			(0.144)		
PPOL	-1.135^{***}	$(-1.777,\ -0.493)$	34	-0.326	$(-0.668,\ 0.016)$	6	-0.185	$(-0.467,\ 0.096)$	1
	(0.328)			(0.175)			(0.144)		
PDS	-1.058^{***}	$(-1.189,\ -0.928)$	41	-0.642^{***}	$(-0.843,\ -0.441)$	6	-0.631^{***}	$(-0.892,\ -0.370)$	1
	(0.067)			(0.102)			(0.133)		

Note: $N = 88$. $\hat{\beta}$ is the coefficient; CI is the 95% confidence interval; $\dot{\boldsymbol{z}}$ reports the *sum* of selected covariates across Stage 1 and Stage 2 (for PSL, one stage only). Standard errors appear in parentheses below coefficients. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 3: LASSO Estimation Results with Polynomial Feature Expansion

	CV			**BRT**			**BCCH**		
	$\hat{\beta}$	CI	$\dot{z}$	$\hat{\beta}$	CI	$\dot{z}$	$\hat{\beta}$	CI	$\dot{z}$
Post-Lasso Method									
PSL	0.228^{**}	$(0.022,\ 0.435)$	25	-0.270^{**}	$(-0.482,\ -0.058)$	3	-0.121	$(-0.419,\ 0.177)$	0
	(0.106)			(0.108)			(0.152)		
PDL	-1.212^{***}	$(-1.830,\ -0.593)$	61	-0.341	$(-0.701,\ 0.019)$	5	-0.121	$(-0.357,\ 0.115)$	0
	(0.316)			(0.184)			(0.121)		
PPOL	-1.002^{***}	$(-1.571,\ -0.433)$	58	-0.341	$(-0.701,\ 0.019)$	5	-0.121	$(-0.357,\ 0.115)$	0
	(0.290)			(0.184)			(0.121)		
PDS	-0.844^{***}	$(-0.927,\ -0.761)$	61	-0.628^{***}	$(-0.824,\ -0.432)$	5	-0.121	$(-0.419,\ 0.177)$	0
	(0.042)			(0.100)			(0.152)		

Note: $N = 88$. $\hat{\beta}$ is the coefficient; CI is the 95% confidence interval; $\dot{z}$ reports the *sum* of selected covariates across Stage 1 and Stage 2 (PSL has one stage). Standard errors in parentheses below coefficients. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

References

Barro, R. (1989). *Economic Growth in a Cross Section of Countries*. en. Tech. rep. w3120. Cambridge, MA: National Bureau of Economic Research, w3120. DOI: 10.3386/w3120.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). "Inference on Treatment Effects after Selection among High-Dimensional Controls". en. In: *The Review of Economic Studies* 81.2, pp. 608–650. ISSN: 0034-6527, 1467-937X. DOI: 10.1093/restud/rdt044.

Belloni, A. et al. (2012). "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain". en. In: *Econometrica* 80.6, pp. 2369–2429. ISSN: 0012-9682. DOI: 10.3982/ECTA9626.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). "Simultaneous analysis of Lasso and Dantzig selector". In: *The Annals of Statistics* 37.4. ISSN: 0090-5364. DOI: 10.1214/08-AOS620.

Chetverikov, D., Z. Liao, and V. Chernozhukov (2020). *On cross-validated Lasso in high dimensions*. arXiv:1605.02214 [math]. DOI: 10.48550/arXiv.1605.02214.

Sala-I-Martin, X. X. (1997). "I Just Ran Two Million Regressions". In: *The American Economic Review* 87.2. Publisher: American Economic Association, pp. 178–183. ISSN: 00028282.