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

By Astrid Sophie Fugleholm, Jacob Strabo and Jørgen Høst

Characters: Approx 11,217 (including spaces and math).

1 Introduction

Can developing countries expect to catch up to developed countries in terms of economic growth? Or does it require them to make institutional, cultural, or structural changes to their economies? It has been the agenda of many economic researchers to pin-point the determinants of economic growth for centuries. The insights produced play a crucial role in core assumptions of economic models on which world-wide economic policy is based. For the world economy as a whole to grow at its full potential, it is relevant for both developing and developed countries to know whether we can count on the former experiencing catch-up growth. In this paper, we thus evaluate the hypothesis of conditional convergence in a high-dimensional setting.

2 Data

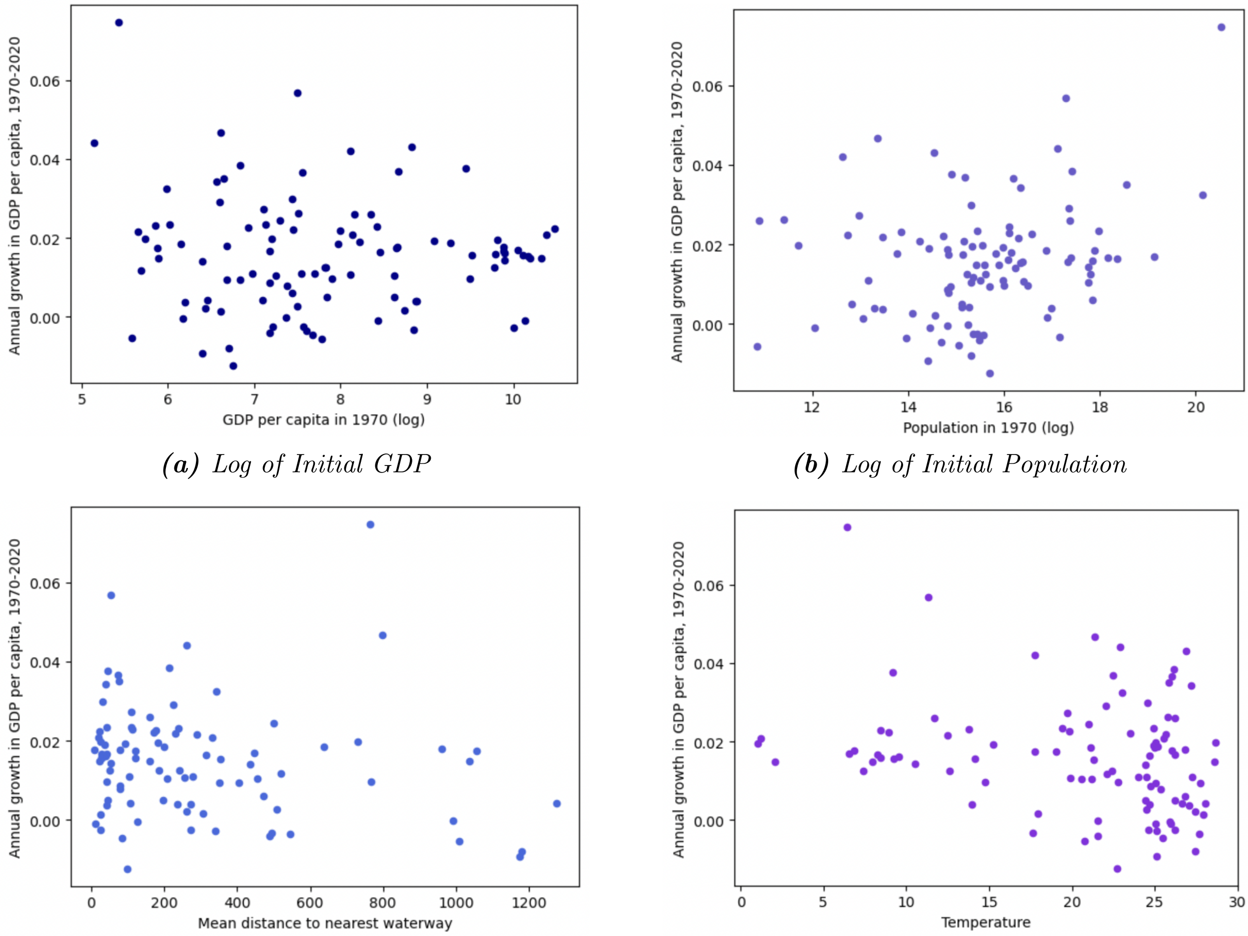
We use cross section data for $N = 214$ countries with observations on 85 characteristics regarding geography, resources, genetic diversity, institutions, history, religion, danger, and education. Our outcome variable is the annual growth rate of GDP per capita, which is only observed for $N = 102$ of the countries. As the aim of our analysis is to examine the relationship between growth and initial level of GDP, the log of initial GDP per capita is our main variable of interest.

Figure 1 show correlations between the growth rate in GDP per capita on the one hand and initial GDP, distance to nearest waterway, initial population, and temperature on the other. It suggests that countries with a lower initial level of GDP, larger initial population, easy access to water, and cold weather experience higher GDP growth. Hence, factors related to the size of the workforce and geography are, in addition to initial GDP, important determinants of GDP growth and may even drive the negative correlation between initial level and growth rate of GDP.

Figure 2 show heat maps of the annual growth rate of GDP, log of initial GDP, annual population growth rate, and investment rate in countries in our sample. *Figures 2a-2b* illustrate that countries with a high initial level of GDP (e.g., the US, Norway, and Australia) relative to others (e.g., China and India) experienced a lower annual growth in GDP, and vice versa. In addition, *figure 2c-2d* demonstrate that a high investment rate and low population growth are important drivers of growth in some countries (e.g., China and India).

In sum, *figure 1-2* show that factors related to workforce size, geography, and investments are, in addition to initial GDP, important determinants of GDP growth. Geographical characteristics likely impact GDP in 1970 and thus drive its negative correlation with GDP growth. Hence, it is not ex ante clear that countries with a lower initial GDP will experience a higher GDP growth.

Figure 1: *Correlations between GDP growth, initial GDP, initial population, distance to waterway, and temperature*



(a) *Log of Initial GDP* ***(b)*** *Log of Initial Population*

(c) *Distance to Nearest Coast or River (km)* ***(d)*** *Temperature (Degrees Celsius)*

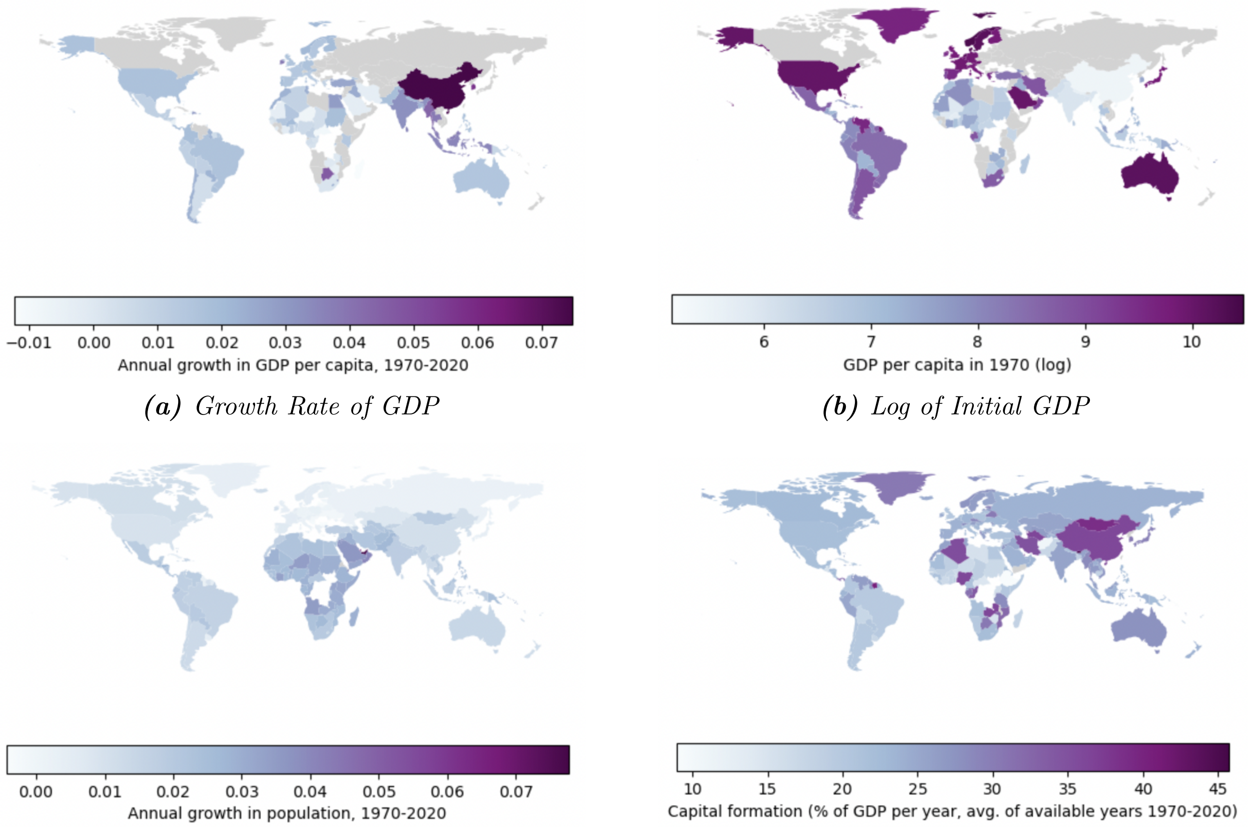
3 Economic Growth

The hypotheses of conditional convergence is that, in the long run, all countries with similar characteristics end up on the same growth path implying that those with an initially low level of GDP experience a higher growth rate of GDP, and vice versa. The relation between growth rate and initial level of GDP can be explored following Barro (1991) in setting up a regression for country i

$$g_i = \beta y_{i0} + \mathbf{z}_i\gamma + u_i, \quad E[u_i|\beta,\gamma] = 0, \tag{1}$$

where g_i is the growth rate of GDP; y_{i0} the log of initial level of GDP per capita; $\mathbf{z}_i$ a vector of control variables; and u_i the idiosyncratic error term capturing all other unobserved factors affecting g_i. Our parameter of interest is β capturing the effect of initial level on growth rate of GDP with $\beta < 0$ implying a negative relation between initial level and growth rate of GDP (β-convergence). Testing the hypothesis

Figure 2: *Heat Maps of Growth in GDP, Initial Level of GDP, Population Growth, and Investment Rate*



(a) *Growth Rate of GDP* ***(b)*** *Log of Initial GDP*

(c) *Population Growth Rate* ***(d)*** *Investment Rate*

of conditional convergence thus amounts to testing the null and alternative hypotheses

$$\mathcal{H}_0 : \beta = 0 \quad \text{and} \quad \mathcal{H}_A : \beta < 0 \tag{2}$$

using a one-sided t-test, which is asymptotically normally distributed under the null. To estimate β in (1), we must choose a set of controls, $\mathbf{z}_i$. To do this, we take an agnostic approach in including potential control variables. First, we discard those variables that have less than 95% of GDP growth observed. Second, we add interactions and squared terms to investigate non-linear effects. This leaves us with a dataset of $N = 92$ countries and $p = 405$ regressors.

4 Econometric Model

When choosing a suitable model to estimate the growth rate, a natural starting point is the OLS estimator.

$$\hat{\beta} = \text{argmin} \frac{1}{n} \sum_{i=1}^{n} (g_i - \mathbf{X_i}' b)^2 \tag{3}$$

Where $\mathbf{X_i} = (y_{i0}, \mathbf{z_i})$ is a composite matrix of our treatment variable and candidate regressors. A key component of the OLS estimator is full invertibility, which breaks down whenever we have more candidate regressors p than observations n. Further, when p is large, as in this case, the prediction error of OLS is also expected to be high, why OLS is not necessarily the best fit for this analysis. This can be seen when taking the expected average prediction error:

$$E\left[\frac{1}{N}\sum_{i=1}^{n}(\mathbf{X}_i'\hat{\beta} - \mathbf{X}_i'\beta)^2\right] = \frac{\sigma^2 p}{n} \tag{4}$$

So to do inference when $p > n$, we need a method of 'shutting down' potentially unimportant regressors, and find the best fitted linear model to explain growth. One such is LASSO

$$\hat{\beta} = \text{argmin}\frac{1}{n}\sum_{i=1}^{n}((g_i - \mathbf{X_i}'b)^2 + \lambda||b||_1 \tag{5}$$

The minimization problem is similar to that of OLS, except LASSO adds the absolute value of the coefficients, $||b||_1$, multiplied by a penalty term, λ, to improve the generalization of the model. The LASSO estimator seeks to find a global minimum by changing all the coefficients in the model, including setting some coefficients to zero. In this way, LASSO excludes irrelevant, independent variables, and leaves only the most useful explanatory variables - given the end-goal is prediction (Belloni et al. (2014a)). Going further, we can approach this "naïvely" by applying LASSO to (1) and excluding β from the LASSO penalty. We can try to draw inference between β and any remaining regressors in $\mathbf{z_i}$. However, not only does this method rely on perfect model selection, thus potentially inducing omitted variable bias, we also do not know the asymptotic distribution of $\hat{\beta}$ in this high dimensional regime. Therefore, we consider the Post Double LASSO next.

4.1 Post Double LASSO (PDL)

The method of Post Double Lasso solves the issue of unknown asymptotic distribution of $\hat{\beta}$. Consider equation (2) with following augmented 'first-stage' regression:

$$y_{i0} = \mathbf{z_i}\phi + \epsilon_i \tag{6}$$

From (1) and (6), we require $E[u_i|\beta, \gamma] = 0$ and $E[\epsilon_i|\phi] = 0$, resp. Combining these, we have a set of moment conditions that implies $E[(y_{i0} - \mathbf{z_i}\phi)(g_i - \beta y_{i0} - \mathbf{z_i}\gamma)] = 0$. Rearranging yields

$$\check{\beta} = \frac{E[(y_{i0} - \mathbf{z_i}\phi)(g_i - \mathbf{z_i}\gamma)]}{E[(y_{i0} - \mathbf{z_i}\phi)y_{i0}]} \tag{7}$$

$\check{\beta}$ follows an asymptotic distribution $\frac{\sqrt{n}(\check{\beta}-\beta)}{\sigma} \xrightarrow{d}$N(0,1) for $n \to \infty$ with $\sigma^2 := \frac{E[u^2\epsilon^2]}{(E[\epsilon^2])^2}$ and confidence intervals $\check{C}I_{1-\xi} = \left[\check{\beta} \pm q_{1-\xi/2}\frac{\check{\sigma}}{\sqrt{n}}\right]$ based on ξ significance level for the inverse standard normal CDF q, cf. Belloni et al. (2014b). This allows us to draw inference on β without requiring $p < n$. Mechanically, it implies a 'first stage' and a 'second stage' where we LASSO first our outcome variable g_i and second our treatment variable y_{i0} on candidate regressors $\mathbf{z_i}$ with suitable penalty levels for each stage.

4.2 Estimating the penalty term

There are several ways of estimating the penalty term λ. We use the Bickel-Ritov-Tsybakov Rule (BRT) and the Belloni-Chen-Chernozhukov-Hansen Rule (BCCH) due to their feasibility for out-of-sample predictions.

4.2.1 The Bickel-Ritov-Tsybakov Rule (BRT)

To estimate λ, we first choose α and c and then estimate λ as

$$\hat{\lambda}^{BRT} := \frac{2c\sigma}{\sqrt{n}}\phi^{-1}\left(1 - \frac{\alpha}{2p}\right)\max\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2}, \tag{8}$$

where ϕ is the standard normal CDF. The computation relies on two assumptions: 1) Homoskedasticity, such that ϵ is independent of X, and 2) variance σ^2 of ϵ is known. It may be that these assumptions are not fulfilled, so as an alternative consider the following penalty level.

4.2.2 The Belloni-Chen-Chernozhukov-Hansen Rule (BCCH)

Similar to BRT, we first choose α and c and then estimate λ as

$$\hat{\lambda}^{BRT} := \frac{2c}{\sqrt{n}}\phi^{-1}\left(1 - \frac{\alpha}{2p}\right)\max\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i'(\hat{\beta}(\hat{\lambda}^{pilot})))^2 x_{ij}^2} \tag{9}$$

with

$$\hat{\lambda}^{pilot} := \frac{2c\sigma}{\sqrt{n}}\phi^{-1}\left(1 - \frac{\alpha}{2p}\right)\max\sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 x_{ij}^2} \tag{10}$$

BCCH allows heteroscedastic noise and does not assume that we know the variance of ϵ.

5 Results

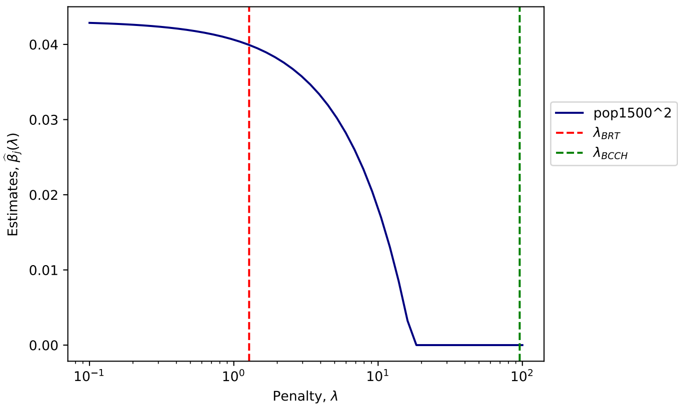
The results from our LASSO estimations are reported in *table 1*. Of things to note, we cannot reject our null of conditional convergence using the Post Double Lasso. For illustration, we included the naïve estimate and its LASSO path. Further, we see that compared to the naïve approach, the penalty level is somewhat lower when using BRT PDL but remarkably higher when using BCCH PDL. BCCH will put a larger penalty on and shrink more coefficients to zero than BRT, so we end up with a model with fewer and more relevant regressors.

Table 1: *Beta-convergence*

	Penalty	β	se	CI	λ_{nave}	λ_{yz}	λ_{bz}	N	p
PDL	`BRT`	-0.20	0.01	-0.22; -0.18	.	1.03	0.97	92	405
PDL	`BCCH`	-0.14	0.01	-0.16; -0.12	.	17.35	7.63	92	405
Naïve	`BRT`	0.17	0.02	0.2; 0.14	1.28	.	.	92	405

Note: Standard 5% significance levels apply. Naïve describes the approach outlined above, where we use (Single) LASSO to select control variables and then apply OLS. Only the squared level of population in 1500 (`pop1500`) survived this process. λ_{yz} and λ_{bz} denotes the penalty levels used in the Post Double LASSO (PDL) for the 'first' and 'second' stage.

Figure 3: *Naïve LASSO Path*



6 Discussion

The opposing conclusions on β-convergence from the naïve and PDL approaches is due to the former's focus on prediction. The naïve LASSO includes the set of regressors which collectively maximize predictive power, discarding regressors' individual ability to predict GDP growth. In a pair of two correlated, strong predictors, it will exclude one as it adds little to the predictive power provided by the other. This means, that we potentially remove causal explanatory variables, and keep the ones with the best fit for the model. In contrast, the two-stage procedure of the PDL ensures that both regressors are included. It is therefore likely that PDL has kept relevant regressors that the naïve LASSO has excluded, and thereby provide a more reliable estimate of β. Indeed, as *figure 3* shows that the naïve LASSO shuts down all regressors except the squared population size in 1500 (`pop1500`), and given *figure 1a*, this can explain the positive β estimate.

Importantly, high-dimensional methods like LASSO are not a panacea to omitted variable bias. Recall that for our results to hold, we require that $E[u_i|\beta, \gamma] = 0$ and $E[\epsilon_i|\phi] = 0$, i.e. that the idiosyncratic errors are uncorrelated with our candidate regressors. The assumptions will be violated if factors such as international finance, trade intensity, and corruption, which are not included in our $\mathbf{z}_i$, also affect GDP growth. In that case, our β estimates are biased and results not useful.

7 Conclusion

In this paper, we investigate the hypothesis of conditional convergence. Given that we are in a high-dimensional setting with more observations on candidate regressors than countries, we cannot use OLS and instead turn to Post Double LASSO. We find support for the hypothesis of conditional convergence, implying that countries with a lower initial GDP will experience a higher GDP growth. Our results rely on the assumption that our set of candidate regressors include all relevant factors. If this does not hold, our β estimate will be biased due to omitted variables. However, as our final data set includes 405 regressors, any omitted variable bias is presumably small.

References

Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics*, 106(2):407–443.

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