Advanced Microeconometrics

Kaya Maria Friis Jensen & Anna Høybye Schmidt

High-Dimensional Linear Models and Convergence in Economic Growth

ku-da.pdf

ku-farve.pdf

Project 2

Autumn 2024

Contents

1 Introduction

The theory of convergence is a central topic in growth theory, addressing whether initially poor countries grow faster than initially rich countries. Economies are said to converge to a common growth path if the correlation between initial GDP per capita and growth in GDP per capita is negative. We investigate the convergence theory by implementing a high-dimensional sparse model setting, which allows for a large set of regressors with a small sample of countries. We use the Post-Double LASSO (PDL) estimator and find no evidence that initially poorer countries grow faster than initially richer ones.

2 Theory

To investigate whether initially poor countries tend to grow faster than initially rich countries, we follow [Barro, 1991], and regress the average annual growth rate of GDP per capita (g_i) in country i on the log of initial GDP per capita (y_{i0})

$$g_i = \beta y_{i0} + \mathbf{z}_i\gamma + u_i, \quad \mathbb{E}[u_i|y_{i0}, \mathbf{z}_i] = 0, \quad i = 1, ..., n, \tag{2.1}$$

where $\mathbf{z}_i$ is a vector of control variables and u_i is the unobservable error term. We are interested in exploring β, which encompasses the effect of initial GDP per capita on growth in GDP per capita when controlling for $\mathbf{z}_i$. If $\beta < 0$ then convergence occurs.

Since the number of countries (n) is relatively low while the number of regressors (p) is relatively high, model 2.1 is high-dimensional. In a high-dimensional setting, OLS might not be a good predictor as its prediction error is

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(\hat{g}_i{}^{LS} - g_i)^2\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i'\hat{\beta}^{LS} - X_i'\beta)^2\right] = \frac{\sigma^2 p}{n} \tag{2.2}$$

where $\hat{g}_i{}^{LS} = X_i'\hat{\beta}^{LS}$ is the LS predictor and $g_i = X_i'\beta$ is the *optimal* predictor. As p is large and n is small $\frac{p}{n} \not\to 0$, i.e. the prediction error of OLS will not approach zero. This makes OLS inadequate for the analysis, if we assume sparsity. We propose the Least Absolute Shrinkage and Selection Operator (LASSO) estimator as a solution.

2.1 LASSO Estimator

The LASSO estimator is similar to the OLS estimator. However, in addition, the LASSO estimator imposes sparsity by putting a penalty on regressors through the sum of the absolute values of the coefficients

$$\hat{\beta}(\lambda) \in \arg\min_{b} \left\{ \frac{1}{n} \sum_{i=1}^{b} (g_i - \beta y_{i0} - \mathbf{z}_i' \gamma)^2 + \lambda ||b_j|| \right\}, \quad \lambda \geq 0 \tag{2.3}$$

where the first term is the LS minimization problem, and the second term is the penalty term, where $||b_j|| = |\beta| + \sum_{j=1}^{p} |\gamma_j|$ and λ is the penalty level. The estimator depends on the assumption of sparsity, i.e. that only a subset of coefficients are different from zero. The size of the penalty term λ determines which coefficients are selected into the model, facing a trade-off between model fit and regularization. In the following section, we elaborate on the selection of an appropriate penalty term.

2.2 Post Double Lasso

The LASSO estimator uses sparsity to enhance the explanatory power of the model. As such, regressors that correlate with initial GDP per capita may be discarded if they do not increase the model's predictive power. This can lead to omitted variable bias which motivates using a Post Double Lasso (PDL) estimator to estimate Model 2.1. Thus, the augmented model, in reduced form relation between the treatment and controls, is $y_{i0} = \psi \mathbf{z'}_i + v_i, \quad E[v_i|\mathbf{z}_i] = 0$ where $E[v_i|\mathbf{z}_i]$ and the exogeneity condition from Eq. 2.2 comprise the conditional independence assumption from which we can set up the moment condition $E[(y_0 - \mathbf{z}'\psi_o)(g - \beta_0 y_0 - \mathbf{z'}\gamma_0)]$, which allows us to identify the true parameter (β_0) as

$$\beta_0 = \frac{E[(y_0 - \mathbf{z}'\psi_o)(g - \beta_0 y_0 - \mathbf{z}'\gamma_0)]}{E[(y_0 - \mathbf{z}'\psi_o)y_0]}$$

As such, we use the LASSO estimator to estimate $\mathbf{z}$ and y_0 from g, which allows us to find $\hat{\beta}$ and $\hat{\gamma}$. Next, we estimate $\mathbf{z}$ from y_0 which we then use to find $\hat{\psi}$. From, this we can apply the analogy principle to obtain

$$\hat{\beta}^{PDL} = \frac{\sum_{i=1}^{n}(y_{i0} - \mathbf{z}_i\hat{\psi})(g_i - \mathbf{z}_i\hat{\gamma})}{\sum_{i=1}^{n}(y_{i0} - \mathbf{z}_i\hat{\psi})y_{i0}} \tag{2.4}$$

The LASSO estimator does not have an asymptotic distribution. However, the PDL estimator converges to a normal distribution under some conditions:

$$\sqrt{n}(\hat{\beta}^{PDL} - \beta_0) \xrightarrow{D} N(0, \sigma^2), \quad \sigma^2 = \frac{\mathbb{E}(v^2 w^2)}{\left(\mathbb{E}(w^2)\right)^2}$$

By applying the analogy principle again, we estimate the variance (σ^2) as follows

$$\sigma^2 = \frac{n^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i^2\hat{v}_i^2}{\left(n^{-1}\sum_{i=1}^{n}\hat{v}_i^2\right)^2} \tag{2.5}$$

Lastly, we consider the theory of convergence by testing whether countries with a low initial level of GDP per capita tend to grow faster than those with a high initial level. Specifically, we test null hypothesis of no convergence and alternative hypothesis of convergence: $H_0 : \beta_0 = 0, H_A : \beta_0 < 0$. As the alternative hypothesis is one-sided and strictly bounded from below, the t-test's critical value is (-1.65) on a 5 pct. significance level.

2.3 Selection of Penalty Term

To select the penalty term, we consider three methods. The first method for selecting the penalty term is Cross-Validation (CV). However, we choose not to consider CV because [Belloni et al., 2014] argues that it is inappropriate for inference for a specific coefficient, which is relevant to our project. Second is the Bickel-Ritov-Tsybakov (BRT) rule, which relies on two conditions to produce a valid penalty term: first, conditional homoskedasticity, and second, the variance of the residuals is known. These conditions ensure that the variance of the residuals is independent of the regressors and that the penalty can be calculated. However, we do not know the variance in our case, and homoskedasticity may not hold. We instead rely on the Belloni-Chen-Chernozhukov-Hansen rule (BCCH) which does not require conditional homoskedasticity nor knowledge about the variance of the residuals. Usually, this makes the BCCH penalty larger as the rule is more restrictive.

we calculate the penalty term in three steps. For the first step, we choose the probability tolerance $\alpha \in (0,1)$, which is set to 0.05 and the markup $c > 1$, which we set to 1.1 as in [Chetverikov and Sørensen, 2023]. Second, we run the pilot LASSO

$$\hat{\lambda}^{\text{pilot}} := \frac{2c}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2p}\right) \max_{1 \le j \le p} \sqrt{\frac{1}{n}\sum_{i=1}^{n}(g_i - \bar{g})^2 X_{ij}^2} \tag{2.6}$$

In the third step, we use $\hat{\lambda}^{\text{pilot}}$ to calculate the residuals: $\hat{\epsilon}_i = g_i - \bar{g} = g_i - \mathbf{X}_i' \hat{\beta}^{\text{pilot}}\left(\hat{\lambda}^{\text{pilot}}\right)$, that goes into the estimation of the BCCH penalty term:

$$\hat{\lambda}^{\text{BCCH}} := \frac{2c}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2p}\right) \max_{1 \le j \le p} \sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^2 X_{ij}^2} \tag{2.7}$$

3 Data and Choice of Controls

We consider a cross-sectional dataset with 214 countries. Within the dataset, we treat GDP per capita growth as the outcome variable and the log of initial GDP per capita as the explanatory variable of interest. We limit the dataset to those, where data for GDP per capita growth is available, bringing the sample down to 102 countries. To avoid omitted variable bias within the model, we aim to include relevant controls for economic growth over time. Drawing from existing literature, we choose to include controls that belong to one of the following three categories: **institutions**, [Robinson and Acemoglu, 2012], [Acemoglu et al., 2005], **geography**, [Diamond, 1999], and **education**, [Cuaresma, 2006]. We also include the investment rate and population growth, [Galor, 2012], such that we end up with 29 possible controls, stated in Table 1.[1] We decide to exclude countries with missing observations which reduces our sample to 71 countries. The sample is fairly balanced across continents, cf. Figure 1. We construct first-order interaction terms and get 428 controls when excluding interaction terms with zero variance, as these remain constant across countries and thus do not contribute to explaining differences. Lastly, we standardize all variables by subtracting the mean and dividing by the standard deviation.

[1] Initially, we have 30 possible controls. However, we discover that the variables *demCGV* and *demBMR* are linearly dependent which causes issues with the rank condition. Therefore, we choose to remove the variable *demBMR* implying that we end up with 29 controls.

4 Analysis

Results of Eq. 2.3 are displayed in Table 2. Model (1) and (2) are OLS estimators, (1) has no controls, (2) has 29 control. We have some non-zero controls for the BRT penalty models (3) and (4), but as argued in the theory section, we don't expect the penalty term to be valid. Model (5) with 29 controls and (6) with the interaction terms for all controls, use the BCCH-penalty term, which leads to one non-zero control in Model (5) and no non-zero controls in Model (6). Thus, the BCCH-penalty in Model (6) is so high that no controls have non-zero coefficients, implying that $\hat{\beta}^{BCCH}$ is identical to $\hat{\beta}^{OLS}$ indicating that none of the controls have significant explanatory power. Model (5) and (6) predict that the point estimate of $\hat{\beta}$ is respectively (-0.1783) and (-0.1720), indicating that a one standard deviation increase in the log of initial GDP per capita *may* decreases subsequent annual average growth by approximately 0.17 standard deviations. However, as both point estimates are insignificant we cannot reject the null hypothesis of no convergence.

5 Discussion and Conclusion

In this paper, we have investigated the theory of convergence using a high-dimensional model. We don't find any implication that initially poor countries grow faster than initially rich countries, based on our included regressors.

Our conclusion is, furthermore, based on a relatively small sample (71 countries), which may limit the statistical significance of our estimates.

Our parameter estimates may suffer omitted variable bias. For example, [Clark, 2007] argues that economic growth depends on cultural factors, which is just one of the groups of variables that we do not consider. We decided to exclude cultural variables, based on compelling examples in [Robinson and Acemoglu, 2012] explaining how local regions can have different growth paths if they belong to other countries, meaning different institutions despite having the same cultural background. Also, sample selection bias may affect our parameter estimates, for instance, due to the exclusion of countries with missing values. If missing values are correlated with initial GDP per capita and subsequent growth, our estimates may be biased. Our estimates may also be biased if countries with low initial GDP per capita and low subsequent growth are underrepresented in our sample, as this could bias the results toward rejecting the hypothesis of no convergence. However, we are not too concerned about this as Africa is well represented in our sample, cf. figure 1.

6 References

[Acemoglu et al., 2005] Acemoglu, D., Johnson, S., and Robinson, J. A. (2005). *Institutions as a Fundamental Cause of Long-Run Growth*, volume 1 of *Handbook of Economic Growth*. Elsevier.

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

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

[Chetverikov and Sørensen, 2023] Chetverikov, D. and Sørensen, J. R.-V. (2023). *Selecting Penalty Parameters of High-Dimensional M-Estimators using Bootstrapping after Cross-Validation*. arXiv:2104.04716 [math.ST].

[Clark, 2007] Clark, G. (2007). *A farewell to alms: a brief economic history of the world*. Princeton University Press.

[Cuaresma, 2006] Cuaresma, J. C. (2006). Convergence of educational attainment levels in the oecd: More data, more problems? *Economics of Education Review*, Volume 25, 2(2):173–178.

[Diamond, 1999] Diamond, J. M. (1999). *Guns, germs, and steel*, volume 521. Books on Tape New York.

[Galor, 2012] Galor, O. (2012). The demographic transition: causes and consequences. *Cliometrica*, 6(1):1–28.

[Robinson and Acemoglu, 2012] Robinson, J. A. and Acemoglu, D. (2012). *Why nations fail: The origins of power, prosperity and poverty*. Profile London.

7 Appendix

Table 1: List of Included Control Variables, Distributed on Categories

Institutions	*marketref, dem, demCGV, demreg*
Geography	*tropicar, distr, distcr, suitavg, temp, suitgini, elevavg, elevstd, kgatr, precip, area, abslat, cenlong, area_ar, rough, landlock, africa, asia, ocenia, americas*
Education	*ls_bl, lh_bl*
Miscellaneous	*investment_rate, pop_growth*

Table 2: Estimation Results for Six Different Model Specifications

	(1) OLS	(2) OLS	(3) PDL	(4) PDL	(5) PDL	(6) PDL
Initial log (gdp)	-0.172	-0.0094	-0.2562	-0.198	-0.1783	-0.172
se	0.1218	0.0078	0.2059	0.1798	0.148	0.1441
Controls	0	29	29	428	29	428
Non-zero controls	0	29	3	4	1	0
Observations	71	71	71	71	71	71
Penalty term	-	-	BRT	BRT	BCCH	BCCH
λ_{dz}	-	-	0.5726	0.7039	0.9104	1.1921
λ_{yx}	-	-	0.5896	0.7227	1.2197	2.679
t-statistic	-1.4122	-1.2051	-1.2443	-1.1012	-1.2047	-1.1936

Figure 1: Number of Countries in the Sample, Distributed by Continent

bar plot.png