Wiener Institut für Internationale Wirtschaftsvergleiche

The Vienna Institute for International Economic Studies

Working Papers | 55 | June 2009

Ulrike Schneider and Martin Wagner

Catching Growth Determinants with the Adaptive Lasso

wiiw Working Papers published since 2002:

No. 55 U. Schneider and M. Wagner: Catching Growth Determinants with the Adaptive Lasso. June 2009

No. 54 J. Crespo-Cuaresma, N. Foster and R. Stehrer: The Determinants of Regional Economic Growth by Quantile. May 2009

No. 53 C. Lennon: Trade in Services and Trade in Goods: Differences and Complementarities. April 2009

No. 52 J. F. Francois and C. R. Shiells: Dynamic Factor Price Equalization and International Convergence. March 2009

No. 51 P. Esposito and R. Stehrer: Effects of High-Tech Capital, FDI and Outsourcing on Demand for Skills in West and East. March 2009

No. 50 C. Fillat-Castejón, J. F. Francois and J. Wörz: Cross-Border Trade and FDI in Services. February 2009

No. 49 L. Podkaminer: Real Convergence and Inflation: Long-Term Tendency vs. Short-Term Performance. December 2008

No. 48 C. Bellak, M. Leibrecht and R. Stehrer: The Role of Public Policy in Closing Foreign Direct Investment Gaps: An Empirical Analysis. October 2008

No. 47 N. Foster and R. Stehrer: Sectoral Productivity, Density and Agglomeration in the Wider Europe. September 2008

No. 46 A. Iara: Skill Diffusion by Temporary Migration? Returns to Western European Work Experience in Central and East European Countries. July 2008

No. 45 K. Laski: Do Increased Private Saving Rates Spur Economic Growth? September 2007

No. 44 R. C. Feenstra: Globalization and Its Impact on Labour. July 2007

No. 43 P. Esposito and R. Stehrer: The Sector Bias of Skill-biased Technical Change and the Rising Skill Premium in Transition Economies. May 2007

No. 42 A. Bhaduri: On the Dynamics of Profit- and Wage-led Growth. March 2007

No. 41 M. Landesmann and R. Stehrer: Goodwin's Structural Economic Dynamics: Modelling Schumpeterian and Keynesian Insights. October 2006

No. 40 E. Christie and M. Holzner: What Explains Tax Evasion? An Empirical Assessment based on European Data. June 2006

No. 39 R. Römisch and M. Leibrecht: An Alternative Formulation of the Devereux-Griffith Effective Average Tax Rates for International Investment. May 2006

No. 38 C. F. Castejón and J. Wörz: Good or Bad? The Influence of FDI on Output Growth. An industry-level analysis. April 2006

No. 37 J. Francois and J. Wörz: Rags in the High Rent District: The Evolution of Quota Rents in Textiles and Clothing. January 2006

No. 36 N. Foster and R. Stehrer: Modelling GDP in CEECs Using Smooth Transitions. December 2005

No. 35 R. Stehrer: The Effects of Factor- and Sector-biased Technical Change Revisited. September 2005

No. 34 V. Astrov, Sectoral Productivity, Demand, and Terms of Trade: What Drives the Real Appreciation of the East European Currencies? April 2005

No. 33 K. Laski: Macroeconomics versus 'Common Sense'. December 2004

No. 32 A. Hildebrandt and J. Wörz: Determinants of Industrial Location Patterns in CEECs. November 2004

No. 31 M. Landesmann and R. Stehrer: Income Distribution, Technical Change and the Dynamics of International Economic Integration. September 2004

No. 30 R. Stehrer: Can Trade Explain the Sector Bias of Skill-biased Technical Change? May 2004

No. 29 U. Dulleck, N. Foster, R. Stehrer and J. Wörz: Dimensions of Quality Upgrading in CEECs. April 2004

No. 28 L. Podkaminer: Assessing the Demand for Food in Europe by the Year 2010. March 2004

No. 27 M. Landesmann and R. Stehrer: Modelling International Economic Integration: Patterns of Catching-up, Foreign Direct Investment and Migration Flows. March 2004

No. 26 M. Landesmann and R. Stehrer: Global Growth Processes: Technology Diffusion, Catching-up and Effective Demand. January 2004

No. 25 J. Wörz: Skill Intensity in Foreign Trade and Economic Growth. November 2003; revised version January 2004

No. 24 E. Christie: Foreign Direct investment in Southeast Europe: a Gravity Model Approach. March 2003

No. 23 R. Stehrer and J. Wörz: Industrial Diversity, Trade Patterns and Productivity Convergence. November 2002; revised version July 2003

No. 22 M. Landesmann and R. Stehrer: Technical Change, Effective Demand and Economic Growth. April 2002

No. 21 E. Christie: Potential Trade in South-East Europe: A Gravity Model Approach. March 2002

Ulrike Schneider is Junior Professor at the Institute for Mathematical Stochastics, University of Göttingen. Martin Wagner is Researcher at the Institute for Advanced Studies, Vienna.

Financial support from the Jubiläumsfonds of the Oesterreichische Nationalbank (Project number: 11688) and the European Commission's DG Regional Policy within the study 'Analysis of the Main Factors of Regional Growth: An in-depth study of the best and worst performing European regions' (contract no. 2007.CE.16.0.AT.209) coordinated by the Vienna Institute for International Economic Studies (wiiw) is gratefully acknowledged. The comments of participants of a workshop held at wiiw, of participants of the 3rd International Conference on Computational and Financial Econometrics in Limassol, of seminar participants at Universitat Pompeu Fabra in Barcelona and the University of Gießen, and of M. Hashem Pesaran are gratefully acknowledged. The usual disclaimer applies.

Ulrike Schneider and
Martin Wagner

Catching Growth Determinants with the Adaptive Lasso

Contents

Abstract

This paper uses the adaptive Lasso estimator to determine the variables important for economic growth. The adaptive Lasso estimator is a computationally very simple procedure that can perform at the same time model selection and consistent parameter estimation. The methodology is applied to three data sets, the data used in Sala-i-Martin et al. (2004), in Fernandez et al. (2001) and a data set for the regions in the European Union. The results for the former two data sets are similar in several respects to those found in the published papers, yet are obtained at a negligible fraction of computational cost. Furthermore, the results for the European regional data highlight the importance of human capital for economic growth.

Keywords: *adaptive Lasso, economic convergence, growth regressions, model selection*

JEL classification: *C31, C52, O11, O18, O47*

1 Introduction

The econometric analysis of economic growth and potential economic convergence of per capita GDP has been a major research topic in the last decades. This highly active field of research has been revived among others by the influential contributions of Baumol (1986), Barro (1991) and Barro and Sala-i-Martin (1992). Numerous different econometric approaches and techniques have been employed, as surveyed by Durlauf et al. (2005). Yet, few definite results have emerged, in the words of Durlauf et al. (2005, p. 558):

"The empirical study of economic growth occupies a position that is notably uneasy. Understanding the wealth of nations is one of the oldest and most important research agendas in the entire discipline. At the same time, it is also one of the areas in which genuine progress seems hardest to achieve. The contributions of individual papers can often appear slender. Even when the study of growth is viewed in terms of a collective endeavor, the various papers cannot easily be distilled into a consensus that would meet standards of evidence routinely applied in other fields of economics."

The largest part of the empirical studies undertaken deals with so-called growth (or Barro) regressions, in which the average growth rate of per capita GDP is regressed on initial per capita GDP and a potentially large set of additional explanatory variables. Such equations have their original motivation in first order approximations (around the steady state) of the Solow-Swan or Ramsey-Cass-Koopmans versions of the one-sector growth model, as illustrated in Barro (1991) or Mankiw et al. (1992). Based on these approximations numerous researchers have estimated vast amounts of equations including a large variety of additional explanatory variables. Due to the relatively weak link between the specified equations and growth theory such empirical studies have to a certain extent to be seen as data mining exercises.

Given the data mining character of growth regressions many empirical strategies have been followed to separate the wheat from the chaff. Sala-i-Martin (1997b) runs two million regressions and uses a modification of the extreme bounds test of Leamer (1985), used in the growth context earlier also by Levine and Renelt (1992), to single out what he calls 'significant' variables. Fernandez et al. (2001) and Sala-i-Martin et al. (2004) use Bayesian model averaging (BMA) techniques to identify important growth determinants. Doing so necessitates the estimation of a large number of potentially ill-behaved regressions (e.g. in case of near multi-collinearity of the potentially many included regressors). Typically it is impossible, due to the sheer number of possible models, to get exact BMA estimates and therefore only some approximate estimates based on Markov chain Monte Carlo methods are computed. Clearly, also the specification of priors in the context of growth regressions is a delicate issue given that little prior information is available concerning either the relevance of individual variables or the number of variables relevant. In this respect the work of Magnus et al. (2009) is interesting in that it provides a computationally simple approach to BMA that is based on specific priors with a clear interpretation. Hendry and Krolzig (2004) use, similar to Hoover and Perez (2004), a general-to-specific modelling strategy to cope with the large amount of regressors while avoiding the estimation of a large number of equations. Clearly, also in a general-to-specific analysis a certain number of regressions, typically greater than one, has to be estimated.

In this paper we determine the variables important for economic growth by resorting to recently developed statistical techniques designed to achieve at the same time model selection and consistent parameter estimation. In particular we use the so-called *adaptive Lasso* (Least Absolute Shrinkage and Selection Operator) estimator, a variant of the Lasso estimator (Tibshirani, 1996), proposed by Zou (2006) which we briefly describe in Section 2. The adaptive Lasso estimator is an example of a *penalized least squares estimator*. Due to their strong computational benefits, these types of estimators have received a lot of attention in the recent statistical literature, but have not yet been applied within the context of economic growth.

The approach we use here has several advantages. First, as mentioned above, it is computationally very cheap. Using e.g. the algorithm proposed by Efron et al. (2004), the entire sequence of regressions which implicitly considers all submodels roughly has the same computational cost as just one single OLS regression including all regressors. Second, the procedure can generally handle ill-behaved regressions or the case where there are more explanatory variables than observations. This is an advantage compared to general-to-specific approaches. Third, the version of the estimator employed in this paper is scale independent, see Section 2, i.e. it adapts to changes of units in the variables just like the OLS estimator. This is an advantage of the adaptive LASSO estimator over and above many other related estimation procedures currently investigated in the statistics literature, such as the Lasso estimator, the SCAD estimator (Fan and Li, 2001), or the elastic net (Zou and Hastie, 2005) which do not share this property. Finally, for those who prefer to use classical statistical methods over Bayesian methods and estimates based on a single model over averages, the adaptive Lasso provides exactly that.

We apply the method to several data sets, where two of them are well-known data sets taken from widely cited papers. Note also that, compared to a typical BMA analysis, no prior choice concerning model size has to be made but the model size is itself an outcome of the procedure. The first data set is the data presented in Sala-i-Martin et al. (2004), containing 67 explanatory variables for 88 countries. The second one is the data set investigated in Fernandez et al. (2001), which is in turn based on data used in Sala-i-Martin (1997b) and contains 41 explanatory variables for 72 countries. The third data set we use comprises the 255 European NUTS2 regions in the 27 member states of the European Union and contains 48 explanatory variables. The results for the first two data sets are in several aspects similar to the findings in the original papers. For both data sets the adaptive Lasso estimator selects slightly less than 15 explanatory variables, with about 10 of them coinciding with the most important ones of the original papers (measured there by posterior inclusion probabilities). All coefficient estimates have the expected signs and plausible values. The results for the regional data (for which fewer core economic data are available to act as explanatory variables) indicate the importance of human capital for economic growth, proxied by medium and high education.

This paper is organized as follows. In Section 2 we describe the statistical methods that we apply. Section 3 contains the empirical analysis and results, and Section 4 briefly summarizes and concludes. Two appendices follow the main text. Appendix A describes the regional data set and Appendix B presents the results of the adaptive LASSO estimation sequence graphically.

2 The Adaptive Lasso Estimator

The adaptive Lasso estimator, a variant of the Lasso estimator, is a special case of the general class of penalized least squares (PLS) estimators. For a linear regression model $y = X\beta + \varepsilon$ ($y, \varepsilon \in \mathbb{R}^N$, $X \in \mathbb{R}^{N\times k}$, $\beta \in \mathbb{R}^k$), a PLS estimator $\hat{\beta}$ of β is defined as the solution of the minimization problem

$$\min_{\beta\in\mathbb{R}^k} \|y - X\beta\|^2 + \lambda_N \; pen(\beta), \tag{1}$$

where $pen(\beta)$ is the *penalty function* and λ_N the so-called *tuning parameter*. Clearly, different types of penalty terms give rise to different estimators. The general class of Bridge estimators was introduced by Frank and Friedman (1993) and refers to estimators defined by (1) with $pen(\beta) = \sum_{j=1}^{k} |\beta_j|^{\gamma}$. Note that $\gamma = 2$ corresponds to the well-known Ridge estimator. For $\gamma \leq 1$, due to the structure of the underlying optimization problem, coordinates of the estimated coefficient vector $\hat{\beta}$ can (potentially) be exactly equal to zero and in that sense the resulting estimator can be viewed to also perform model selection. The case $\gamma = 1$ where the penalty term is given by the l_1-norm of the coefficient vector corresponds to the Lasso estimator. This estimator was treated separately in Tibshirani (1996) who also introduced the name Lasso. The adaptive Lasso estimator, as introduced in Zou (2006), has a randomly weighted l_1 penalty function defined by $pen(\beta) = \sum_{j=1}^{k} |\beta_j|/|\tilde{\beta}_j|^{\vartheta}$, where $\tilde{\beta}$ is any $\sqrt{N}$-consistent 'initial' estimator of β, typically the OLS estimator of the full model (if available). We stick to this choice of $\tilde{\beta}$ together with using $\vartheta = 1$, since then the estimator becomes *scale-independent* in the sense that replacing a regressor x_j $(1 \leq j \leq k)$ by a scalar multiple, say cx_j with $c \in \mathbb{R}$, will result in an estimator where the corresponding component $\hat{\beta}_j$ is now replaced by $\frac{1}{c}\hat{\beta}_j$.

The asymptotic properties of the adaptive Lasso estimator (and PLS estimators in general) mainly depend on the choice of the tuning parameters λ_N as N goes to infinity. Under standard assumptions on the regression model (iid errors ε_i with mean zero and finite variance; $X'X/N \to C$ as $N \to \infty$ for some positive definite matrix C), the following holds for the adaptive Lasso estimator. If λ_N converges to a finite number, the estimator is *tuned conservatively*, i.e. it performs *conservative model selection*, finding the correct zeros with probability less to one but choosing only correct models asymptotically. If $\lambda_N \to \infty$ as $N \to \infty$ and $\lambda_N/N \to 0$, the estimator is *tuned consistently*, i.e. it performs *consistent model selection*, finding the correct zeros with probability equal to one. While the latter regime seems to be preferable, it should be noted that consistently tuned estimators are not without problems since they generally exhibit inferior distributional properties compared the corresponding conservatively tuned or unpenalized estimator. Among those issues are unbounded estimator risk (Leeb and Pötscher, 2008) and large confidence intervals (Pötscher and Schneider, 2008).

We now discuss computational issues of the adaptive Lasso estimator. Solutions to the minimization problem defining this estimator can be computed very efficiently exploiting the specific structure of the problem. It can be shown that the components of the solution to the corresponding optimization problem are piecewise linear in the tuning parameter λ_N, see e.g. Rosset and Zhu (2007). Exploiting this property, the estimator can easily be computed for *all* tuning parameters $\lambda_N \in [0, \infty)$, leading to so-called (piecewise linear) solution paths for the coefficients corresponding to each variable. These solution paths are initiated at λ_N equal to infinity where all coefficients are equal to zero and ensued up to λ_N equal to zero corresponding to the OLS estimator (in case it is uniquely defined). In each step along this sequence, one variable is either included or removed from the current 'active' subset, i.e. the set containing the variables whose coefficients are not equal to zero in that step, as illustrated in Figure 1 and Figure 2 in Appendix B for the Sala-i-Martin et al. (2004) data set. These figures should be read as follows. First note that on the horizontal axis λ_N is drawn in *de*creasing order, running from ∞ to 0, which is standard for these plots in the literature. On the horizontal axis the steps of the sequence are plotted equidistantly, e.g. in Figure 1 the first twenty steps are plotted and in between each of these steps the estimated coefficients are linear in λ_N. This implies that here the scaling of the horizontal axis is in general not linear with respect to the tuning parameter λ_N. In the example of Figure 1 the value of the tuning parameter at which the first estimated coefficient starts to become non-zero is $\lambda_N = 0.51$. The numbers on the right hand side of the graph indicate the variable number, e.g. the index

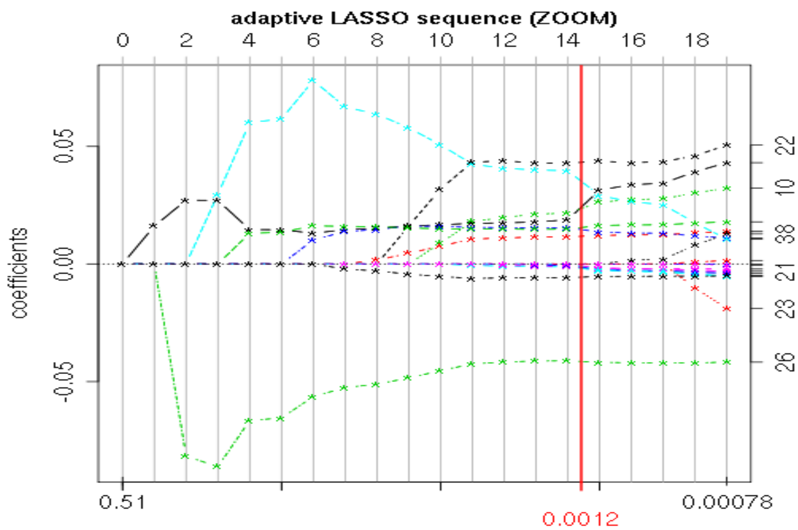

Figure 1: Coefficient paths of adaptive Lasso estimation for the Sala-i-Martin et al. (2004) data set for the first 20 steps of the adaptive Lasso estimation sequence (zoom of Figure 2). The vertical line at $\lambda_N = 0.0012$ indicates the optimal tuning parameter λ_N chosen by cross-validation. See Section 2 for an explanation of the plot.

26 refers to the share of expenditure of government consumption of GDP in 1961 (for the variable abbreviations and their numbers for the Sala-i-Martin et al. (2004) data see Table 1). The (dash-dotted) line corresponding to such an index plots the coefficient of this variable as a function of the tuning parameter. The vertical line at $\lambda_N = 0.0012$ indicates the optimal choice of the tuning parameter according to cross-validation, see below.

To provide a 'final' subset of variables together with a corresponding estimate of the parameter vector different approaches are used to choose the tuning parameter, with the most common one given by cross-validation. This procedure is known to potentially lead to conservative model selection, i.e. it may lead to the inclusion of some variables whose true coefficients are equal to zero, see e.g. Leng et al. (2006). Given the favorable results obtained by cross-validation in a variety of experiments and the fact that conservatively tuned estimators possess more desirable distributional properties, as mentioned above, the results presented in this paper are based on cross-validation. We provide the value of λ_N selected this way for each data set in the corresponding plots, yet we emphasize that this is merely a technical parameter which cannot be compared for different data sets and does not carry further information with regard to the application at hand.

3 Empirical Analysis

As mentioned in the introduction, the empirical analysis is performed for three different data sets. These are the data sets used in Sala-i-Martin et al. (2004), in Fernandez et al. (2001) and a data set covering the 255 NUTS2 regions of the European Union. In the discussion below we retain the variable names from the data files we received from Gernot Doppelhofer for the Sala-i-Martin et al. (2004) data and also use the original names used in the file downloaded from the homepage of the Journal of Applied Econometrics for the Fernandez et al. (2001) data to facilitate the comparison with the results in these papers.

3.1 Sala-i-Martin, Doppelhofer and Miller Data

The data set considered in Sala-i-Martin et al. (2004) contains 67 explanatory variables for 88 countries. The variables and their sources are described in detail in Table 1 in Sala-i-Martin et al. (2004, p. 820–821). The dependent variable is the average annual growth rate of real per capita GDP over the period 1960–1996. In Table 1 we present the sequence of adaptive Lasso moves (i.e. the sequence of variables in- respectively excluded from the set of active variables as $\lambda_N \to 0$) for this data set. As already discussed in the previous section, graphical information concerning the whole sequence of estimated coefficients as a function of the tuning parameter λ_N is presented in Figure 2 in Appendix B and Figure 1 in Section 2.

The full regressor matrix comprising the constant and all 67 explanatory variables is almost multi-collinear, with a reciprocal condition number of 9.38×10^{-20}. The full OLS estimator is thus very imprecisely defined at best. To be precise, in order to invert the $X'X$ matrix the numerical tolerance has to be set to an extremely small number. To avoid using the ill-defined OLS estimator we acknowledge the (numerical) multi-collinearity in the regressor matrix and use as initial estimator the (standard) Lasso estimator at the end of the solution path, i.e. for $\lambda_N = 0$. Due to continuity of the coefficient paths in the tuning parameter, this corresponds to the solution of the normal equations with the smallest l_1-norm. One could also use different regularized initial estimators, e.g. the Ridge estimator, for which, however, the resulting estimator would not be scale-invariant anymore and additionally a choice concerning the Ridge parameter would have to be made. Note that for the two other data sets considered we use the OLS estimator as initial estimator since for those no multi-collinearity problems arise.[1]

Choosing the tuning parameter by cross-validation leads to a model with 15 regressors including the constant. The estimation results are presented in Table 2. The following 14 variables are important to explain cross-country growth, listed in alphabetical order of variable abbreviation, where we also show the sign of the corresponding coefficient in parenthesis: BUDDHA (fraction of population Buddhist in 1960, positive), CONFUC (fraction of population Confucian in 1960, positive), EAST (East Asian dummy, positive), GDE (average share of public expenditure on defense, positive), GDP (log per capita GDP in 1960, negative), GEEREC

[1] It has been suggested to us to also experiment with using as initial estimator the coefficients obtained from the simple regressions of the growth rate of per capita GDP separately on each variable and the constant. This, however, did not lead to sensible results, which could stem from the fact that this initial estimator is not in general consistent.

1 + Const (1)	27 + ECORG (16)	53 + TROPPOP (64)
2 + GVR61 (26)	28 + GDE (20)	54 + LHCPC (34)
3 + GDE (20)	29 + POP65 (51)	55 − SPAIN (60)
4 + P (45)	30 + SIZE (58)	56 + MALFAL (37)
5 + IPRICE (30)	31 − GVR61 (26)	57 + NEWSTATE (40)
6 + EAST (15)	32 + LIFE (35)	58 + SPAIN (60)
7 + TROPICAR (63)	33 + POP (50)	59 + SCOUT (57)
8 + BUDDHA (6)	34 + ABSLATIT (2)	60 + TOTIND (62)
9 + GEEREC (22)	35 + SQPI (47)	61 + PI (46)
10 + CONFUC (10)	36 + CIV (8)	62 + DENSI (13)
11 + LAAM (31)	37 + POP15 (49)	63 + LIFE (35)
12 + MALFAL (37)	38 + EUROPE (18)	64 + YRSOPEN (67)
13 + REVCOUP (55)	39 + LT100CR (36)	65 + OIL (41)
14 + GDP (21)	40 + PRIGHTS (48)	66 + RERD (54)
15 + SAFRICA (56) **	41 + ENGFRAC (17)	67 + WARTIME (65)
16 + MINING (38)	42 + DENS (11)	68 + HERF (28)
17 + OTHFRAC (44)	43 + PRIEXP (52)	69 + WARTORN (66)
18 + GGCFD (23)	44 + PROT (53)	70 + OPEN (42)
19 + SPAIN (60)	45 + GVR61 (26)	71 + ORTH (43)
20 + DENSC (12)	46 + COLONY (9)	72 + LANDAREA (32)
21 + MUSLIM (39)	47 + HINDU (29)	73 + SOCIALIST (59)
22 − GDE (20)	48 − MALFAL (37)	74 + DPOP (14)
23 + GOVSH61 (25)	49 + BRIT (5)	75 + LANDLOCK (33)
24 + H (27)	50 + AIRDIST (3)	76 + CATH (7)
25 + GOVNOM1 (24)	51 − LIFE (35)	77 + TOT1DEC (61)
26 + FERT (19)	52 + ZTROPICS (68)	78 + AVELF (4)

Table 1: Sequence of adaptive Lasso moves for the Sala-i-Martin et al. (2004) data. The entries in the table read as follows: The integers enumerate the step, '+' indicates inclusion of a variable in the corresponding step, whereas '−' refers to exclusion of a variable. The included/excluded variable is referenced by name as well as by its number in the data base, listed in parenthesis in the above table. The set of active variables based on cross-validation is given by the variables included in the active set up to the step indicated by **.

	$\hat{\beta}_{AL}$	sd_{ZOU}	t_{ZOU}	rk($\mathbb{P}_{post}$)	$\mathbb{E}(post)$
Const	0.024139	0.004868	4.959203		
BUDDHA	0.011779	0.003219	3.659490	16	0.002340
CONFUC	0.023542	0.005946	3.959658	9	0.011212
EAST	0.014604	0.002346	6.224814	1	0.017946
GDE	0.035172	0.032333	1.087783	45	0.000952
GDP	−0.000790	0.000386	−2.045769	4	−0.005849
GEEREC	0.043222	0.036496	1.184297	48	0.002720
GVR61	−0.041468	0.015460	−2.682270	18	−0.004594
IPRICE	−0.000069	0.000016	−4.437095	3	−0.000065
LAAM	−0.001902	0.000943	−2.017295	11	−0.001901
MALFAL	−0.001512	0.000465	−3.253112	7	−0.003957
P	0.015468	0.003412	4.532908	2	0.021374
REVCOUP	−0.001382	0.000685	−2.016397	41	−0.000205
SAFRICA	−0.001010	0.000337	−2.995597	10	−0.002265
TROPICAR	−0.005602	0.001578	−3.551155	5	−0.008308

Table 2: Estimation results for the Sala-i-Martin et al. (2004) data. The first three result columns correspond to the adaptive Lasso estimates, with the standard errors and t-values computed as described in Zou (2006). The column labelled rk($\mathbb{P}_{post}$) reports the ranks according to posterior inclusion probabilities from Sala-i-Martin et al. (2004, Table 3, p. 828–829) and the column labelled $\mathbb{E}(post)$ reports the *unconditional* posterior means computed from Sala-i-Martin et al. (2004, Table 4, p. 830) for mean prior model size 7.

(average share of public expenditure on education, positive), GVR61 (share of expenditure on government consumption of GDP in 1961, negative), IPRICE (investment price, negative), LAAM (Latin American dummy, negative), MALFAL (index of malaria prevalence in 1966, negative), P (primary school enrollment rate, positive), REVCOUP (number of revolutions and coups, negative), SAFRICA (sub-Saharan Africa dummy, negative), TROPICAR (fraction of country's land in tropical area, negative). The coefficient signs are all as expected and also the magnitude of the coefficients is plausible and not out of line from other findings in the literature, see also below.

With two exceptions (GDE and GEEREC) the coefficients are statistically different from zero, with the standard errors computed according to Zou (2006) based on Tibshirani (1996). The two variables with insignificant coefficients are both related to government expenditures, namely the average share of public expenditure on defense (GDE) and the average share of public expenditure on education (GEEREC). Out of the government expenditure related variables only the government consumption share of GDP (GVR61) appears to be significant with negative impact on growth. The negative coefficient sign is in line with a stylized relationship in public finance referred to as Wagner's law (formulated by the German economist Adolph Wagner in the nineteenth century), which states that richer countries have a higher public expenditure share. Thus, in case of convergence in which richer countries grow slower, this is consistent with a negative coefficient of the government consumption share of GDP. However, like many empirical growth studies we do not find strong evidence for government expenditure related variables to be major determinants of economic growth. Here we find three government expenditure variables, but two of them with coefficients that are not significantly different from zero.

The fourth results column in Table 2 displays the ranks according to the posterior inclusion probabilities computed from Sala-i-Martin et al. (2004, Table 3, p. 828–829). There is a substantial degree of similarity of our results to theirs in that we find 8 of their top 10 variables (and 9 of their top 14 variables). Of the top 14 variables of Sala-i-Martin et al. (2004) we do not find the following. Population density in coastal areas in the 1960s (DENSC), life expectancy in 1960 (LIFEEXP), fraction of GDP in mining (MINING), the dummy for Spanish colony (SPAIN) and the number of years open (YRSOPEN). Our results suggest instead the importance of the following variables (where we only report the statistically significant variables, which excludes the two mentioned government expenditure variables). The fraction of Buddhists in the population (BUDDHA), the government consumption share of GDP in the 1960s (GVR61) and the number of revolutions and coups (REVCOUP).

In the last column of Table 2 we report the (unconditional with respect to inclusion) posterior means of the estimated coefficients as given in Sala-i-Martin et al. (2004, Table 4, p. 830). The signs of our coefficient estimates coincide with the signs of the posterior means reported in the final column throughout all variables. Some of the posterior means of Sala-i-Martin et al. (2004) are substantially smaller than our estimates. This reflects the fact that, especially for the variables with high inclusion probability ranks, the inclusion probabilities are very small. This in turn leads to small unconditional posterior means, which clearly reflects the shrinkage character of model averaging.

3.2 Fernandez, Ley and Steel Data

The data set used by Fernandez et al. (2001) is based on the data set used in Sala-i-Martin (1997b). In particular they select a subset of the Sala-i-Martin data that contains the 25 variables singled out as important by Sala-i-Martin (1997b). These variables are available for 72 countries. To these they add further 16 variables which are also available for these 72 countries, resulting in 41 explanatory variables in total. The dependent variable is the average annual growth rate of real per capita GDP over the period 1960–1992. A detailed description of the variables and their sources is contained in the working paper Sala-i-Martin (1997a, Appendix 1).

Cross-validation (see Table 3 for the sequence of variables included) leads to an equation including 16 explanatory variables counting the intercept. Graphical information concerning the sequence of estimated coefficients as a function of the tuning parameter is given in Appendix B in Figures 3 and 4.

1 + GDPsh560 (4)	16 + Muslim (35)	31 + RFEXDist (10)
2 + LifeExp (3)	17 + NEquipInv (12)	32 + WarDummy (22)
3 − GDPsh560 (4)	18 + LabForce (42) **	33 + Catholic (29)
4 + Confucious (30)	19 + BlMktPm (15)	34 + Rev&Coup (21)
5 + SubSahara (18)	20 + EcoOrg (6)	35 + Foreign (9)
6 + EquipInv (11)	21 + Buddha (28)	36 + Age (26)
7 + LatAmerica (17)	22 + CivlLib (24)	37 + Popg (40)
8 + GDPsh560 (4)	23 + SpanishCol (39)	38 + AbsLat (25)
9 + Const (1)	24 + English (8)	39 + Area (16)
10 + HighEnroll (19)	25 + FrenchCol (32)	40 + YrsOpen (7)
11 + Mining (5)	26 + OutwarOr (14)	41 + std(BMP) (13)
12 + RuleofLaw (38)	27 + BritCol (27)	42 + Jewish (34)
13 + Hindu (33)	28 + Protestants (37)	43 + PolRights (23)
14 + PrScEnroll (2)	29 + PublEdu (20)	44 + Work/Pop (41)
15 + EthnoLFrac (31)	30 + PrExports (36)	

Table 3: Sequence of adaptive Lasso moves for the Fernandez et al. (2001) data set. See caption to Table 1 for further explanations.

The following variables are selected, in alphabetic ordering of variable name: Confucius (share of population Confucian, positive), EquipInv (equipment investment, positive), EthnoLFrac (ethnolinguistic fractionalization, positive), GDPsh560 (log of per capita GDP in 1960, negative), HighEnroll (enrollment rates in higher education, negative), Hindu (share of population Hindu, negative), LabForce (size of labor force, positive), LatAmerica (dummy for Latin America, negative), LifeExp (life expectancy in 1960, positive), Mining (fraction of GDP in mining, positive), Muslim (share of population Muslim, positive), NEquipInv (non-equipment investment, positive), PrScEnroll (primary school enrollment in 1960, positive), RuleofLaw (rule of law, positive) and SubSahara (dummy for sub-Saharan Africa, negative). Again, negative and positive indicate the signs of the corresponding coefficients.

Our results concerning variable selection correspond to a large extent with those of Fernandez

	$\hat{\beta}_{AL}$	sd_{ZOU}	t_{ZOU}	rk($\mathbb{P}_{post}$)
Const	0.059002	0.008513	6.931117	
Confucious	0.056703	0.008732	6.493496	2
EquipInv	0.162682	0.025104	6.480282	4
EthnoLFrac	0.001475	0.000737	2.000760	28
GDPsh560	-0.011876	0.001536	-7.732135	1
HighEnroll	-0.028726	0.012228	-2.349162	34
Hindu	-0.013690	0.005989	-2.285798	19
LabForce	2.6×10^{-8}	1.5×10^{-8}	1.771539	25
LatAmerica	-0.007057	0.001635	-4.317139	13
LifeExp	0.000719	0.000127	5.664690	3
Mining	0.021357	0.006051	3.529288	11
Muslim	0.000978	0.000401	2.436605	6
NEquipInv	0.003127	0.001034	3.023599	12
PrScEnroll	0.003888	0.001623	2.395375	14
RuleofLaw	0.006382	0.001762	3.621853	7
SubSahara	-0.018488	0.002445	-7.562503	5

Table 4: Estimation results for the Fernandez et al. (2001) data. The final column reports the rank according to posterior inclusion probabilities from Fernandez et al. (2001, Table I, p. 569). See caption to Table 2 for further explanations.

et al. (2001, Table I, p. 569) with respect to posterior inclusion probabilities. In particular 11 of the 15 variables included in our results are among the top 15 of the variables of Fernandez et al. (2001). The 4 of their top 15 variables that are not included in our results are, ordered according to decreasing posterior inclusion probability, given by years of openness (YrsOpen), degree of capitalism (EcoOrg), and the fractions of Protestants (Protestants) and of Buddhists (Buddha) in the population. The 4 differing variables that we obtain by applying the adaptive Lasso estimator are the measure of ethnolinguistic fractionalization (EthnoLFrac, positive), enrollment rate in higher education (HighEnroll, negative), fraction of Hindus in the population (Hindu, negative) and size of the labor force (LabForce, positive).

The coefficient for the size of the labor force, a variable that is meant to capture the size of the economy, is significantly different from zero at the 10% level but not at the 5% level. As a side remark note that the small value for the coefficient corresponding to the size of the labor force stems from the fact that Fernandez et al. (2001) use employed persons in their data base, and not e.g. employed persons in thousands or millions. For reasons of comparability we have decided not to change the scaling of this variable. The negative sign of the coefficient corresponding to the high education enrollment rate may merely reflect the fact that countries with a well and broadly functioning higher education system in the 1960s have mainly been rather well-developed rich countries which have subsequently grown below average.

3.3 European Regional Data

The third data set we analyze is a regional data set containing 48 explanatory variables for the 255 NUTS2 regions in the 27 member states of the European Union. The data and variables are described in Appendix A. The dependent variable is the average annual growth rate of per capita GDP over the period 1995–2005. On a regional level it is more difficult to obtain core economic data, hence many of the variables listed in Table 8 in Appendix A are related to infrastructure characteristics (meant in a very broad sense e.g. also including dummy variables whether the regions are located on the seaside or at country borders) and labor market variables (unemployment and activity rates, as well as some broad education characteristics in the working age population). Given that there are both large intra- and inter-country differences in the economic performance of the European regions our set of variables to perform the adaptive Lasso estimation sequence contains country dummies for the 19 out of the 27 countries that consist of more than just one region.

Cross-validation leads to termination of the estimation sequence at step 13, resulting in an

1 + Const (1)	28 + EREL0 (33)	55 − Const (1)
2 + ShSM (44)	29 + gPOP (3)	56 + DUMc10 (56)
3 + ShSH (43)	30 − ARL0 (41)	57 + ShLLL (46)
4 + EREL0 (33)	31 + EREH0 (31)	58 + DUMc26 (67)
5 + GDPCAP0 (2)	32 + URT0 (38)	59 + DUMc20 (62)
6 + DUMc6 (54)	33 + RegPent27 (15)	60 + DUMc15 (61)
7 + ARL0 (41)	34 + RegObj1 (16)	61 + Settl (12)
8 + Capital (17)	35 + Seaports (19)	62 + Dist.de71 (47)
9 − EREL0 (33)	36 + DUMc12 (58)	63 + Temp (28)
10 + AccessRail (26)	37 + ShCE0 (5)	64 + DUMc21 (63)
11 + DUMc27 (68)	38 + DUMc1 (50)	65 + OUTDENS0 (7)
12 + URT0 (38)	39 + URH0 (35)	66 + ARH0 (39)
13 + DUMc14 (60) **	40 + ShJK0 (6)	67 + RoadDens (21)
14 + EMPDENS0 (9)	41 + URL0 (37)	68 + Const (1)
15 + DUMc5 (53)	42 + DUMc11 (57)	69 + DUMc23 (65)
16 + ERET0 (34)	43 + RegCoast (13)	70 + DUMc22 (64)
17 + AccessAir (25)	44 − URH0 (35)	71 + DUMc2 (51)
18 + DUMc9 (55)	45 + DistCap (48)	72 + AccessRoad (27)
19 + URM0 (36)	46 + Hazard (29)	73 + URH0 (35)
20 + ARM0 (40)	47 + INTF (10)	74 − ARH0 (39)
21 + DUMc3 (52)	48 + DUMc13 (59)	75 + RegBoarder (14)
22 + TELF (11)	49 + Airports (18)	76 + ConnectAir (23)
23 + DUMc24 (66)	50 + ART0 (42)	77 + ARH0 (39)
24 − URT0 (38)	51 + HRSTcore (30)	78 + RailDens (22)
25 + shGFCF (49)	52 + ARL0 (41)	79 + EREM0 (32)
26 + POPDENS0 (8)	53 + ShSL (45)	80 + ConnectSea (24)
27 + AirportDens (20)	54 + ShAB0 (4)	

Table 5: Sequence of adaptive Lasso moves for the regional data set including country dummies for all countries consisting of more than one region. See caption to Table 1 for further explanations.

equation with 11 regressors including the intercept, see Tables 5 and 6. Graphical information concerning the sequence of estimated coefficients as a function of the tuning parameter is given in Appendix B in Figures 5 and 6.

The included explanatory variables and the coefficient signs are in alphabetical order: AccessRail (measure of accessibility by railroad, negative), ARL0 (activity rate of low educated in 1995, negative), Capital (dummy for capital city, positive), GDPCAP0 (log of per capita GDP in 1995, negative), ShSH (share of high educated in labor force, positive), ShSM (share of medium educated in labor force, positive) and URT0 (unemployment rate total in 1995, negative). Furthermore, three country dummies are selected: DUMc6 (dummy for Germany, negative), DUMc14 (dummy for Ireland, positive) and DUMc27 (dummy for UK, negative). The coefficient for the activity rate of low educated (ARL0) is significantly different from 0 only at the 10% level and the coefficient for the unemployment rate total (URT0) is not significantly different from 0 even at the 10% level. The signs of the coefficients are (with exception of AccessRail) in line with expectations. With respect to rail accessibility it should be noted that European railroad infrastructure has been built to a very large degree before the sample period of 1995–2005. In particular a large number of the regions that are best accessible by railroads have experienced fast growth and development in much earlier periods than over the sample period and are now slower growing regions with high development levels. Some well-connected regions hosting from today's perspective 'old industries', as e.g. the German Ruhr area, even experience difficulties in the industrial restructuring process in the sample period. These two observations explain the negative coefficient for AccessRail. Clearly, this example again highlights the need for careful interpretation of growth regression results.

The set of human capital and labor market variables included in the specified equation hints

	$\hat{\beta}_{AL}$	sd_{ZOU}	t_{ZOU}
Const	0.151051	0.010037	15.048697
AccessRail	−0.001065	0.000247	−4.304885
ARL0	−0.004462	0.002465	−1.809736
Capital	0.008043	0.000854	9.416480
GDPCAP0	−0.014692	0.001032	−14.235610
ShSH	0.058607	0.007617	7.693931
ShSM	0.016201	0.003677	4.405531
URT0	−0.005236	0.003975	−1.317379
DUMc6	−0.007992	0.001163	−6.870098
DUMc14	0.002599	0.000496	5.237420
DUMc27	−0.002211	0.000578	−3.827470

Table 6: Estimation results for the regional data set including country dummies for all countries consisting of more than one region. See caption to Table 2 for further explanations.

at the importance of a well-educated labor force for economic growth.[2] Of course, a large share of highly educated people in the labor force requires as a complement the presence of sufficiently many workplaces where these skills are required, i.e. enough companies offering a sufficiently large number of jobs demanding medium or high skills and education. Also the negative coefficient for the activity rate of low educated has to be interpreted the same way, i.e. taking into account the complementarity with typically low value added creating activities. Due to the relatively short time span of only 10 years, using initial values of the explanatory variables may not completely resolve these potential endogeneity issues.

It is worth noting that only three country dummies appear to have explanatory power. These cover two poor growth performers Germany and the UK, and the 'Celtic tiger' Ireland. In

[2]It is important to note here again that the available data mainly allow to capture the influence of human capital. There are no variables that measure or at least proxy physical capital and also proxies for technology are essentially absent. The only exception is share of gross fixed capital formation in gross value added in the initial period. Scarcity of data clearly limits any quantitative study of regional growth to date.

this respect it is to a certain extent surprising that none of the country dummies for formerly centrally planned Central and Eastern European (CEE) countries (with more than one region) is selected by the adaptive LASSO procedure.

4 Summary and Conclusions

In this paper we propose to use the adaptive Lasso estimator to determine the variables relevant for explaining economic growth. The adaptive Lasso estimation sequence essentially has the same computational cost as a single OLS regression and simultaneously performs model selection and consistent parameter estimation. Given the large uncertainty concerning potential growth determinants, reflected in large sets of explanatory variables, a consistent or conservative forward selection procedure avoiding the estimation of potentially ill-behaved regressions including large numbers of variables appears to be particularly useful. The proposed classical methodology avoids both the estimation and averaging of large numbers of models using (either a classical or) a Bayesian framework and also avoids pitfalls related to inference in general-to-specific model selection procedures.[3]

The proposed methodology is implemented for three data sets, namely the data used in Sala-i-Martin et al. (2004), in Fernandez et al. (2001) and a data set covering the regions of the European Union member states. The results for the former two well-studied data sets are quite in line with the findings in the original papers. Yet these results are obtained at a negligible fraction of computational cost. For the Sala-i-Martin et al. (2004) data set we find 12 significant explanatory variables (in an equation comprising 14 explanatory variables) using the t-values according to Zou (2006). 9 of these 12 variables are among the top 14 variables

[3]For classical model averaging and subsequent inference in the context of growth regressions see Wagner and Hlouskova (2009).

with respect to posterior inclusion probability as given in Sala-i-Martin et al. (2004). The findings for the Fernandez et al. (2001) data set also exhibit a high degree of similarity with the findings in the original paper. Here we find 15 explanatory variables, with one of them (size of the labor force) not significant at the 5% level. The set of selected variables contains 11 of the top 14 variables of Fernandez et al. (2001), again according to posterior inclusion probability. For both data sets the sets of variables excluded with our approach compared to the findings in the original papers as well as the sets of included significant explanatory variables not found to be important in the original papers are plausible. Furthermore, the findings obtained for the regional data set (for which only a small number of core economic variables is available) hint at the importance of human capital for economic growth. The results obtained for this data set have to be interpreted with some caution due to two reasons. First, many of the variables included are infrastructure and labor market variables since at the regional level many 'typical' macroeconomic variables are not available. Second, the short time span of the data set may imply that some of the variables are endogenous despite the usage of initial values for many variables. This is an issue less pertinent in the other two more long-run data sets.

The findings in this paper strongly indicate that the adaptive Lasso estimator is indeed an estimation and model selection procedure that can be fruitfully employed in the growth regressions context.

Acknowledgements

Financial support from the Jubiläumsfonds of the Oesterreichische Nationalbank (Project number: 11688) and the European Commission's DG Regional Policy within the study "Analysis of the Main Factors of Regional Growth: An in-depth study of the best and worst performing European regions" (contract no. 2007.CE.16.0.AT.209) coordinated by The Vienna Institute for International Economic Studies is gratefully acknowledged. The comments of participants of a workshop held at the Vienna Institute for International Economic Studies, of participants of the 3rd International Conference on Computational and Financial Econometrics in Limassol, of seminar participants at Universitat Pompeu Fabra in Barcelona and the University of Gießen, and of M. Hashem Pesaran are gratefully acknowledged. The usual disclaimer applies.

References

Barro, R.J. (1991). Economic growth in a cross-section of countries. *Quarterly Journal of Economics* **106**, 407–445.

Barro, R.J. and Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy* **100**, 223–251.

Baumol, W.J. (1986). Productivity growth, convergence and welfare: What the long run data show. *American Economic Review* **76**, 1072–1085.

Durlauf, S.N., Johnson, P.A. and Temple, J.R.W. (2005). Growth econometrics. In Aghion, P. and Durlauf, S.N. (eds.) *Handbook of Economic Growth, Vol. 1A*, Ch.8. Elsevier.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics* **32**, 407–499.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

Fernandez, C., Ley, E. and Steel, M.F.J. (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* **16**, 563–576.

Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools (with discussion). *Technometrics* **35**, 109–148.

Hendry, D.F. and Krolzig, H.-M. (2004). We ran one regression. *Oxford Bulletin of Economics and Statistics* **66**, 799–810.

Hoover, K.D. and Perez, S.J. (2004). Truth and robustness in cross-country growth regressions. *Oxford Bulletin of Economics and Statistics* **66**, 765–798.

Leamer, E. E. (1985). Sensitivity analyses would help. *American Economic Review* **57**, 308–313.

Leeb, H. and Pötscher, B. M. (2008). Sparse estimators and the oracle property, or the return of Hodges' estimator. *Journal of Econometrics* **142**, 201–211.

Leng, C., Lin, Y. and Wahba G. (2006). A note on the Lasso and related procedures in model selection. *Statistica Sinica* **16**, 1273–1284.

Levine, R. and Renelt, D. (1992). Sensitivity analysis of cross-country growth regressions. *American Economic Review* **82**, 942–963.

Magnus, J. R., Powell O., and Prüfer, P. (2009). A comparison of two averaging techniques with an application to growth empirics. Forthcoming in *Journal of Econometrics.*

Mankiw, N.G., Romer, D., and Weil, D.N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics* **107**, 407–437.

Pötscher, B. M. and Schneider, U. (2008). Confidence sets based on penalized maximum likelihood estimators. *Manuscript.* arXiv:0806.1652.

Rosset, S. and Zhu, J. (2007). Piecewise linear regularized solution paths, *Annals of Statistics* **35**, 1012–1030.

Sala-i-Martin, X. (1997a). I just ran 4 million regressions. NBER Working Paper No. 6252.

Sala-i-Martin, X. (1997b). I just ran 2 million regressions. *American Economic Review* **87**, 178–183.

Sala-i-Martin, X., Doppelhofer, G. and Miller, R.I. (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review* **94**, 813–835.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society B* **58**, 267–288.

Wagner, M. and Hlouskova, J. (2009). Growth regressions, principal components and frequentist model averaging. Mimeo.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B* **67**, 301–320.

Appendix A: Description of Regional Data Set

In Table 7 we display the 27 EU member states, the abbreviation we use for the countries, as well as the number of NUTS2 regions in each of the countries. The list of variables is described in Table 8. The base year for price indices is 2000. All variables described as 'initial' and whose variable name ends with 0 display 1995 values. For most of the variables for which we report Eurostat as source the variables used here have been constructed by subsequent calculations based on raw data retrieved from Eurostat.

AT	Austria (9)	FI	Finland (5)	MT	Malta (1)
BE	Belgium (11)	FR	France (22)	NL	Netherlands (12)
BG	Bulgaria (6)	GR	Greece (13)	PL	Poland (16)
CV	Cyprus (1)	HU	Hungary (7)	PT	Portugal (5)
CZ	Czech Rep. (8)	IE	Ireland (2)	RO	Romania (8)
DE	Germany (39)	IT	Italy (21)	SE	Sweden (8)
DK	Denmark (1)	LT	Lithuania (1)	SI	Slovenia (1)
EE	Estonia (1)	LU	Luxembourg (1)	SK	Slovak Rep. (4)
ES	Spain (16)	LT	Latvia (1)	UK	United Kingdom (35)

Table 7: Country abbreviations, names and number of NUTS2 regions in brackets.

MADCAP	Annual growth rate of real GDP per capita over 1995–2005 (dependent variable)	Eurostat
GDPCAP0	Initial real GDP per capita (in logs)	Eurostat
gPOP	Growth rate of population	Eurostat
ShAB0	Initial share of NACE A and B (Agriculture) in GVA	Eurostat
ShCE0	Initial share of NACE C to E (Mining, Manufacturing and Energy) in GVA	Eurostat
ShJK0	Initial share of NACE J to K (Business services) in GVA	Eurostat
OUTDENS0	Initial output density	
POPDENS0	Initial population density	
EMPDENS0	Initial employment density	
INTF	Proportion of firms with own website regression	ESPON
TELF	A typology of estimated levels of business telecommunications access and uptake	ESPON
Settl	Settlement structure	ESPON
RegCoast	Coast: dummy	ESPON
RegBorder	Border: dummy	ESPON
RegPent27	Pentagon EU 27 plus 2: dummy if in pentagon (London, Paris, Munich, Milan, Hamburg)	ESPON
RegObj1	‘Objective 1’ regions, i.e. regions situated within objective 1 regions: dummy	ESPON
Capital	Capital cities: dummy whether region host country capital city	
Airports	Number of airports	ESPON
Seaports	Regions with seaports: dummy	ESPON
AirportDens	Airport density: number of airports per km^2	ESPON
RoadDens	Road density: length of road network in km per km^2	ESPON
Name	Description	Source
Name	Description	Source

Continued on next page

RailDens	Rail density: length of railroad network in km per km^2	ESPON
ConnectAir	Connectivity to comm. airports by car of the capital or centroid of region	ESPON
ConnectSea	Connectivity to comm. seaports by car of the capital or centroid of region	ESPON
AccessAir	Potential accessibility air, ESPON space = 100	ESPON
AccessRail	Potential accessibility rail, ESPON space = 100	ESPON
AccessRoad	Potential accessibility road, ESPON space = 100	ESPON
Temp	Extreme temperatures	ESPON
Hazard	Sum of all weighted hazard values	ESPON
HRSTcore	Human resources in science and technology (core)	Eurostat
EREH0	Initial employment rate - high educated	Eurostat
EREM0	Initial employment rate - medium educated	Eurostat
EREL0	Initial employment rate - low educated	Eurostat
ERET0	Initial employment rate - total	Eurostat
URH0	Initial unemployment rate - high educated	Eurostat
URM0	Initial unemployment rate - medium educated	Eurostat
URL0	Initial unemployment rate - low educated	Eurostat
URT0	Initial unemployment rate - total	Eurostat
ARH0	Initial activity rate high educated	Eurostat
ARM0	Initial activity rate medium educated	Eurostat
ARL0	Initial activity rate low educated	Eurostat
ART0	Initial activity rate total	Eurostat
ShSH	Share of high educated in working age population	Eurostat
Name	Description	Source
Name	Description	Source

Continued on next page

ShSM	Share of medium educated in working age population	Eurostat
ShSL	Share of low educated in working age population	Eurostat
ShLLL	Life long learning	Eurostat
Dist_de71	Distance to Frankfurt	
DistCap	Distance to capital city of resp. country	
shGFCF	Share of GFCF in GVA	Cambridge Econometrics

Table 8: Explanatory variables for the empirical analysis of the regional data set.

Appendix B: Graphics

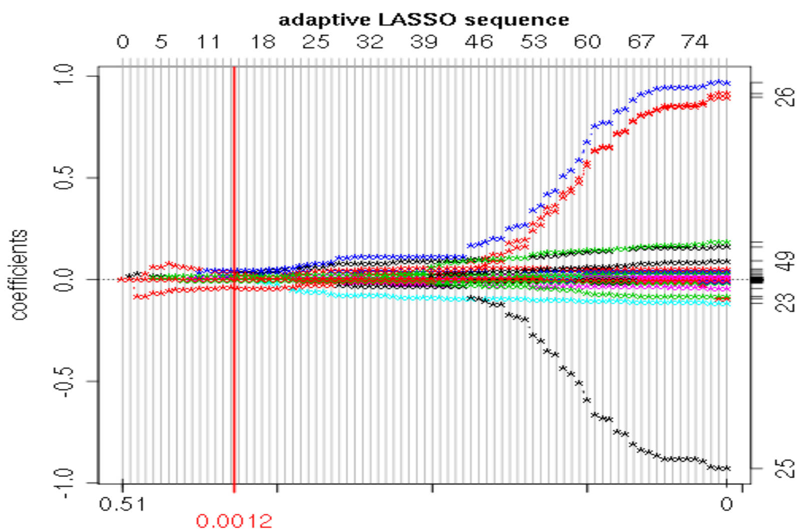

Figure 2: Coefficient paths of adaptive Lasso estimation for the Sala-i-Martin et al. (2004) data set. The vertical line at $\lambda_N = 0.0012$ indicates the optimal tuning parameter λ_N chosen by cross-validation. See Section 2 for an explanation of the plot.

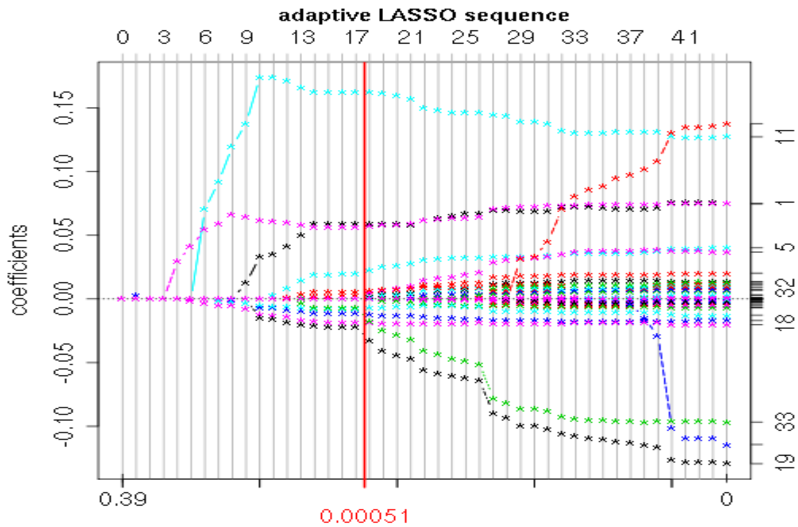

Figure 3: Coefficient paths of adaptive Lasso estimation for the Fernandez et al. (2001) data set. The vertical line at $\lambda_N = 0.00051$ indicates the optimal tuning parameter λ_N chosen by cross-validation. See Section 2 for an explanation of the plot.

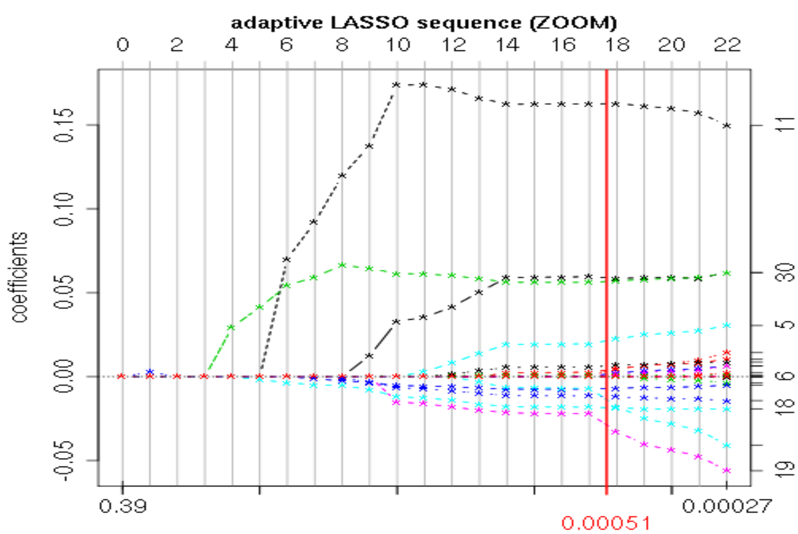

Figure 4: Zoom of Figure 3.

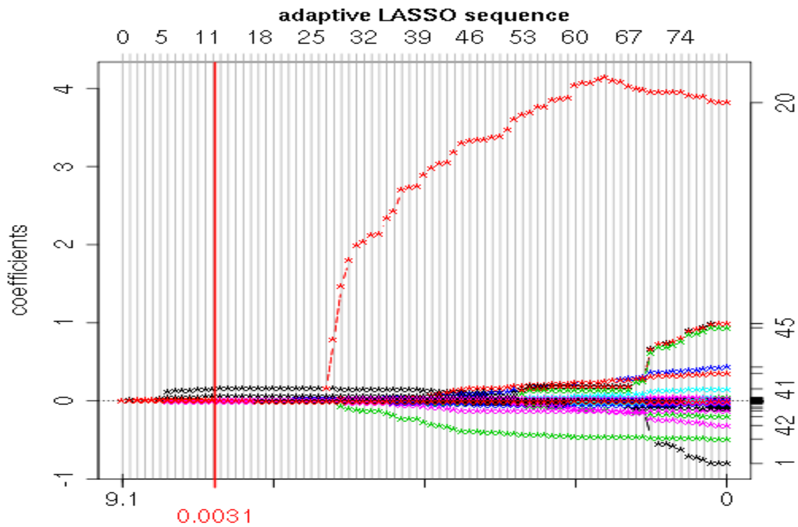

Figure 5: Coefficient paths of adaptive Lasso estimation for the European regional data set. The vertical line at $\lambda_N = 0.0031$ indicates the optimal tuning parameter λ_N chosen by cross-validation. See Section 2 for an explanation of the plot.

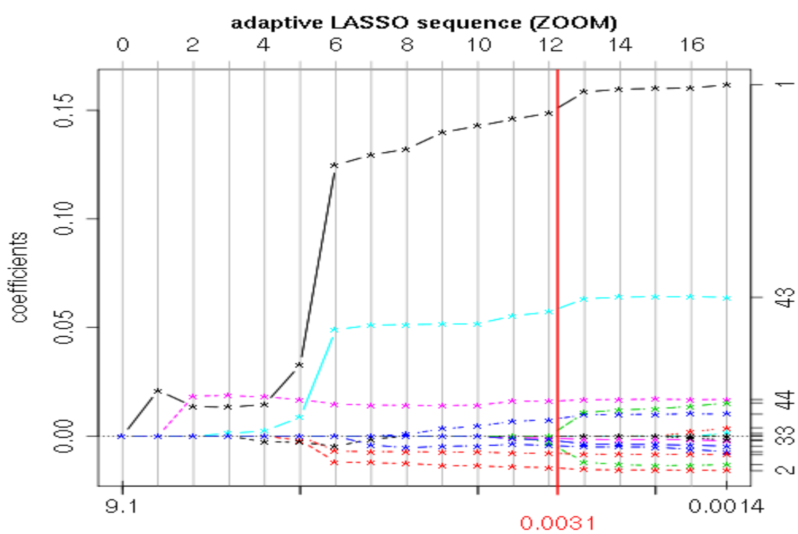

Figure 6: Zoom of Figure 5.

Short list of the most recent wiiw publications (as of June 2009)

For current updates and summaries see also
wiiw's website at www.wiiw.ac.at

Catching Growth Determinants with the Adaptive Lasso

by **Ulrike Schneider and Martin Wagner**

wiiw Working Papers, No. 55, June 2009
34 pages including 8 Tables and 6 Figures
hardcopy: EUR 8.00 (PDF: free download from wiiw's website)

Inequality in Croatia in Comparison

by **Sebastian Leitner and Mario Holzner**

wiiw Research Reports, No. 355, June 2009
38 pages including 6 Tables and 10 Figures
hardcopy: EUR 22.00 (PDF: EUR 20.00)

---

wiiw Monthly Report 6/09

edited by **Leon Podkaminer**

- Crisis management in selected countries of Central, East and Southeast Europe
- The road to China's economic transformation: past, present and future
- Statistical Annex: Selected monthly data on the economic situation in Central and Eastern Europe

wiiw, June 2009
32 pages including 11 Tables and 2 Figures
(exclusively for subscribers to the wiiw Service Package)

The Determinants of Regional Economic Growth by Quantile

by **Jesus Crespo-Cuaresma, Neil Foster and Robert Stehrer**

wiiw Working Papers, No. 54, May 2009
28 pages including 7 Tables and 4 Figures
hardcopy: EUR 8.00 (PDF: free download from wiiw's website)

Changes in the Structure of Employment in the EU and their Implications for Job Quality

by **Robert Stehrer, Terry Ward and Enrique Fernández Macías**

wiiw Research Reports, No. 354, May 2009
106 pages including 29 Tables and 48 Figures
hardcopy: EUR 22.00 (PDF: free download from wiiw's website)

---

wiiw Database on Foreign Direct Investment in Central, East and Southeast Europe, 2009: FDI in the CEECs under the Impact of the Global Crisis: Sharp Declines

by **Gábor Hunya**. Database and layout by **Monika Schwarzhappel**

wiiw Database on Foreign Direct Investment in Central, East and Southeast Europe, May 2009
106 pages including 84 Tables
hardcopy: EUR 70.00 (PDF: EUR 65.00), CD-ROM (including hardcopy): EUR 145.00

MOEL im Sog der Krise

by **Vasily Astrov and Josef Pöschl**

wiiw Research Papers in German language, May 2009
(reprinted from: WIFO-Monatsberichte, Vol. 82, No. 5, May 2009)
14 pages including 6 Tables and 6 Figures
hardcopy: EUR 8.00 (PDF: free download from wiiw's website)

---

wiiw Monthly Report 5/09

edited by **Leon Podkaminer**

- New Hungarian government prescribes bitter medicine
- The steel industry in Central and Eastern Europe: restructuring and prospects
- Transition: unanswered questions
- Statistical Annex: Selected monthly data on the economic situation in Southeast Europe, Russia and Ukraine

wiiw, May 2009
28 pages including 11 Tables
(exclusively for subscribers to the wiiw Service Package)

Trade in Services and Trade in Goods: Differences and Complementarities

by **Carolina Lennon**

wiiw Working Papers, No. 53, April 2009
28 pages including 11 Tables and 3 Figures
hardcopy: EUR 8.00 (PDF: free download from wiiw's website)

---

wiiw Monthly Report 4/09

edited by **Leon Podkaminer**

- Employment and unemployment in the Western Balkans: an assessment
- Skills and export performance
- Financial market regulation and supervision
- Statistical Annex: Selected monthly data on the economic situation in Central and Eastern Europe

wiiw, April 2009
30 pages including 13 Tables and 8 Figures
(exclusively for subscribers to the wiiw Service Package)

Dynamic Factor Price Equalization and International Convergence

by **Joseph F. Francois and Clinton R. Shiells**

wiiw Working Papers, No. 52, March 2009
19 pages including 2 Figures
hardcopy: EUR 8.00 (PDF: free download from wiiw's website)

Effects of High-Tech Capital, FDI and Outsourcing on Demand for Skills in West and East

by **Piero Esposito and Robert Stehrer**

wiiw Working Papers, No. 51, March 2009
21 pages including 6 Tables
hardcopy: EUR 8.00 (PDF: free download from wiiw's website)

---

wiiw Monthly Report 3/09

edited by **Leon Podkaminer**

- Euro or not? Early lessons from the crisis
- Migration from the New to the Old EU Member States: country experiences
- Outsourcing and skills: an empirical investigation
- Statistical Annex: Selected monthly data on the economic situation in Southeast Europe, Russia and Ukraine

wiiw, March 2009

32 pages including 10 Tables and 9 Figures

(exclusively for subscribers to the wiiw Service Package)

Cross-Border Trade and FDI in Services

by **Carmen Fillat-Castejón, Joseph F. Francois and Julia Wörz**

wiiw Working Papers, No. 50, February 2009

49 pages including 15 Tables and 3 Figures

hardcopy: EUR 8.00 (PDF: free download from wiiw's website)

Differentiated Impact of the Global Crisis

by **Vladimir Gligorov, Gábor Hunya, Josef Pöschl et al.**

wiiw Current Analyses and Forecasts. Economic Prospects for Central, East and Southeast Europe, No. 3, February 2009

137 pages including 40 Tables and 16 Figures

hardcopy: EUR 70.00 (PDF: EUR 65.00)

---

wiiw Monthly Report 2/09

edited by **Leon Podkaminer**

- The Republic of Moldova: short-lived recovery
- Patterns of international trade diversification in the transition economies
- Some reflections on the reform of the international financial architecture
- Statistical Annex: Selected monthly data on the economic situation in Central and Eastern Europe (tables)

wiiw, December 2009

34 pages including 12 Tables and 2 Figures

(exclusively for subscribers to the wiiw Service Package)

Entwicklungen der Weltwirtschaft im Kontext der Finanzmarktkrise

by **Michael Landesmann und Roman Stöllinger**

wiiw Research Papers in German language, January 2009

(reprinted from: *Österreichs Außenwirtschaft 2008*, commissioned by the Austrian Federal Ministry of Economics and Labour (BMWA) in the framework of the Research Centre International Economics (FIW) and funded under the internationalization programme 'go international')

22 pages including 6 Figures

hardcopy: EUR 8.00 (PDF: free download from wiiw's website)

---

wiiw Service Package

The Vienna Institute offers to firms and institutions interested in unbiased and up-to-date information on Central, East and Southeast European markets a package of exclusive services and preferential access to its publications and research findings, on the basis of a subscription at an annual fee of EUR 2,000.

This subscription fee entitles to the following package of **Special Services**:

– A free invitation to the Vienna Institute's **Spring Seminar**, a whole-day event at the end of March, devoted to compelling topics in the economic transformation of the Central and East European region (for subscribers to the wiiw Service Package only).

– Copies of, or online access to, ***The Vienna Institute Monthly Report***, a periodical consisting of timely articles summarizing and interpreting the latest economic developments in Central and Eastern Europe and the former Soviet Union. The statistical annex to each *Monthly Report* contains, alternately, country-specific tables or graphs with monthly key economic indicators, economic forecasts, the latest data from the wiiw Industrial Database and excerpts from the wiiw FDI Database. This periodical is not for sale, it can only be obtained in the framework of the wiiw Service Package.

– Free copies of the Institute's ***Research Reports*** (including ***Reprints***), ***Current Analyses and Forecasts****,* ***Country Profiles*** and ***Statistical Reports****.*

– A free copy of the ***wiiw Handbook of Statistics*** (published in October/November each year and containing more than 400 tables and graphs on the economies of Albania, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Montenegro, Poland, Romania, Russia, Serbia, the Slovak Republic, Slovenia and Ukraine)

– Free online access to the **wiiw Monthly Database**, containing more than 1200 leading indicators monitoring the latest key economic developments in ten Central and East European countries.

– **Consulting**. The Vienna Institute is pleased to advise subscribers on questions concerning the East European economies or East-West economic relations if the required background research has already been undertaken by the Institute. We regret we have to charge extra for *ad hoc* research.

– Free access to the Institute's specialized economics library and documentation facilities.

Subscribers who wish to purchase wiiw data sets **on CD-ROM** or special publications not included in the wiiw Service Package are granted considerable **price reductions**.

To
The Vienna Institute
for International Economic Studies
Rahlgasse 3
A-1060 Vienna

- O Please forward more detailed information about the Vienna Institute's Service Package
- O Please forward a complete list of the Vienna Institute's publications to the following address

Please enter me for

- O 1 yearly subscription of *Research Reports* (including *Reprints*) at a price of EUR 180.00 (hardcopy, Europe), EUR 220.00 (hardcopy, overseas) and EUR 140.00 (PDF download with password) respectively
- O 1 yearly subscription of *Current Analyses and Forecasts* a price of EUR 130.00 (hardcopy, Europe), EUR 145.00 (hardcopy, overseas) and EUR 120.00 (PDF download with password) respectively

Please forward

- O the following issue of *Research Reports* ....
- O the following issue of *Current Analyses and Forecasts* ....
- O the following issue of *Working Papers* ....
- O the following issue of *Research Papers in German language* ....
- O the following issue of *wiiw Database on Foreign Direct Investment* ....
- O the following issue of *wiiw Handbook of Statistics* ....
- O (other) ....

....

Name

....

Address

....

Telephone Fax E-mail

....

Date Signature

Herausgeber, Verleger, Eigentümer und Hersteller:
Verein „Wiener Institut für Internationale Wirtschaftsvergleiche“ (wiiw), Wien 6, Rahlgasse 3
Postanschrift: A-1060 Wien, Rahlgasse 3, Tel: [+431] 533 66 10, Telefax: [+431] 533 66 10 50
Internet Homepage: www.wiiw.ac.at

P.b.b. Verlagspostamt 1060 Wien