IHS Economics Series
Working Paper 236
March 2009

Growth Regressions, Principal Components and Frequentist Model Averaging

Martin Wagner
Jaroslava Hlouskova

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Impressum

Author(s):
Martin Wagner, Jaroslava Hlouskova

Title:
Growth Regressions, Principal Components and Frequentist Model Averaging

ISSN: Unspecified

2009 Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
E-Mail: office@ihs.ac.at
Web: www.ihs.ac.at

All IHS Working Papers are available online:
http://irihs.ihs.ac.at/view/ihs_series/
This paper is available for download without charge at:
https://irihs.ihs.ac.at/id/eprint/1907/

236

Reihe Ökonomie

Economics Series

Growth Regressions, Principal Components and Frequentist Model Averaging

Martin Wagner, Jaroslava Hlouskova

236

Reihe Ökonomie

Economics Series

Growth Regressions, Principal Components and Frequentist Model Averaging

Martin Wagner, Jaroslava Hlouskova

March 2009

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Martin Wagner
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-150
email: mawagner@ihs.ac.at

Jaroslava Hlouskova
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
☎: +43/1/599 91-142
email: hlouskov@ihs.ac.at

Abstract

This paper offers two innovations for empirical growth research. First, the paper discusses principal components augmented regressions to take into account all available information in well-behaved regressions. Second, the paper proposes a frequentist model averaging framework as an alternative to Bayesian model averaging approaches. The proposed methodology is applied to three data sets, including the Sala-i-Martin et al. (2004) and Fernandez et al. (2001) data as well as a data set of the European Union member states' regions. Key economic variables are found to be significantly related to economic growth. The findings highlight the relevance of the proposed methodology for empirical economic growth research.

Keywords

Frequentist model averaging, growth regressions, principal components

JEL Classification

C31, C52, O11, O18, O47

Comments
Financial support from the Jubiläumsfonds of the Oesterreichische Nationalbank (Project number: 11688) is gratefully acknowledged. The regional data set has been compiled and made available to us by Robert Stehrer from the Vienna Institute for International Economic Studies (WIIW). The comments of seminar participants at Universitat Pompeu Fabra, Barcelona, are gratefully acknowledged. The usual disclaimer applies.

Contents

1 Introduction

This paper offers a twofold contribution to the empirical growth literature. First, it advocates the use of well-behaved principal components augmented regressions (PCAR) to capture and condition on the relevant information in the typically large set of variables available. Second, it proposes frequentist model averaging as an alternative to Bayesian model averaging approaches commonly used in the growth regressions literature. Model averaging, Bayesian or frequentist, becomes computationally cheap when combined with principal components augmentation.

The empirical analysis of economic growth is one of the areas of economics in which progress seems to be hardest to achieve (see e.g. Durlauf et al., 2005) and where few definite results are established. Large sets of potentially relevant candidate variables have been used in empirical analysis to capture what Brock and Durlauf (2001) refer to as theory open endedness of economic growth, and numerous econometric techniques have been used 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 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) or regressions which suffer from omitted variables biases in case important explanatory variables are not included. 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 a situation in which the potential relevance of large sets of variables is unclear ex ante, any regression including only few explanatory variables is potentially suffering from large biases in case that some or many relevant explanatory variables have been excluded from the regression. However, with the large numbers of variables available it is often infeasible or even impossible to include all variables in the regression. As an example, for one of the data sets employed in this paper, originally used in Sala-i-Martin et al. (2004), the reciprocal condition number of the full regressor matrix including all available explanatory variables is 9.38×10^{-20}. Thus, in the full regression not only is there a very small number of degrees of freedom left, but the coefficients are estimated with additional imprecision due to the numerical (almost) singularity. Consequently, a trade-off has to be made between parsimony of the regression (to achieve low variance but potentially high bias) and the inclusion of as many potentially relevant variables as possible (to achieve low bias at the price of potentially high variance).

In many applications it is conceivable that the researcher has a set of variables in mind whose effect she wants to study in particular. This choice of variables can e.g. be motivated by a specific theoretical model or also by the quest of understanding the contribution to growth of certain factors like human capital related variables. In such a situation, conditional upon an ex ante classification in core and auxiliary variables, the use of principal components augmented regressions allows to focus on untangling the effect of the core variables on growth whilst *controlling for* by *conditioning on* the effects of the auxiliary variables. Performing regression analysis including the core variables and principal components computed from the auxiliary variables allows to take into account 'most' of the information contained in all variables. In particular including principal components of the auxiliary variables in addition to the core variables implies that the bias of the resulting regression will be low, since 'most' of the information contained in all available variables is taken into account (see the discussion below in Section 2). Also, a PCAR typically does not suffer from large estimation variance when the number of core variables and (mutually orthogonal) principal components is reasonably small and multi-collinearity is absent.[1] The coefficients to the core variables in a PCAR measure the effect on growth of each of these variables when considered jointly whilst in addition conditioning out the information contained in the principal components and are in this sense 'robust' estimates.

The fact that in a PCAR most of the information of all variables is included, potentially mitigates the necessity to account for model uncertainty via model averaging. Clearly, however, PCAR analysis can be combined with model averaging, either Bayesian or frequentist (classical). Given the separation in core and auxiliary variables, a natural approach to model averaging is to compute sub-models only with respect to the core variables, whilst including the principal components in all regressions. This has several advantages: First, including

[1] As long as the set of core variables are not multi-collinear, multi-collinearity can be controlled by choosing the number of included principal components accordingly.

in each regression the principal components minimizes potential (omitted variables) biases compared to regressions including only small numbers of variables. Up to now in empirical growth analysis model averaging has been performed mainly over small models, e.g. in the Bayesian model averaging approach pursued in Sala-i-Martin et al. (2004) the mean prior model size is 7 for most of the discussion. This may result in the presence of substantial biases. Second, resorting to PCAR reduces the number of sub-models enormously. If one were to estimate all sub-models (in case all of them can be estimated) for all k variables, then 2^k regressions are necessary, which amounts to 2^{67} regressions for the Sala-i-Martin et al. (2004) data and 2^{41} regressions for the Fernandez et al. (2001) data also considered in this paper. Clearly, these numbers are way too large to estimate all sub-model regressions. The Bayesian literature tries to overcome this limitation by resorting to MCMC sampling schemes designed to approximate the posterior densities of all coefficients. The posteriors depend by definition upon the priors, where as mentioned, large weights are typically put on small models with potentially large biases. In PCAR analysis, the number of regressions to be computed to estimate all sub-models is reduced from 2^k to 2^{k_1}, where k_1 denotes the number of core variables, which typically (at least in our applications) is a rather small number that allows for the estimation of all sub-models. For a Bayesian approach this implies that the posterior distributions can be evaluated exactly, be it analytically or numerically. Furthermore, each of the estimated sub-model PCARs has comparably small omitted variables bias due to the inclusion of principal components.

In this paper we perform model averaging in a frequentist framework, using recent advantages in the statistics literature which allow to perform valid classical inference in a model averaging context, see in particular Claeskens and Hjort (2008) and the brief description in Appendix C. In our analysis we consider four different weighting schemes. One, as a benchmark, uses equal weights for each model and the three others are based on weights derived from information criteria computed for the individual models. These are Mallows model averaging (MMA) advocated by Hansen (2007), and smoothed AIC and smoothed BIC weights considered by Buckland et al. (1997) and studied in detail also in Claeskens and Hjort (2008). Furthermore, we introduce frequentist analogs to quantities considered to be informative in a Bayesian model averaging framework. E.g. we introduce, for any given weighting scheme, the so-called inclusion weight as the classical counterpart of the Bayesian posterior inclusion probability of a variable. Similarly, we consider the distribution of model weights over model sizes.

We apply the methodology to three data sets, with two of them taken from widely cited papers. The first data set is that of Sala-i-Martin et al. (2004), containing 67 explanatory variables for 88 countries. The second one is the Fernandez et al. (2001) data set, based in turn on data used in Sala-i-Martin (1997b), which contains 41 explanatory variables for 72 countries. The third data set comprises the 255 NUTS2 regions of the 27 member states of the European union and contains 48 explanatory variables. These data sets have also been analyzed in Schneider and Wagner (2008), who use the adaptive LASSO estimator of Zou (2006), to perform at the same time model (i.e. variable) selection and parameter estimation. For the two well studied data sets, the sets of variables selected in Schneider and Wagner (2008) correspond closely to the sets of variables found important (measured by posterior inclusion probabilities) in the original papers. To illustrate the PCAR and frequentist model averaging (FMA) approaches advocated in this paper we consider for each of the three data sets the variables selected in Schneider and Wagner (2008) as core variables and all remaining variables as auxiliary variables. The main finding is that our approach singles out, both when considering single PCAR estimates as well as model average estimates, core economic variables as important in explaining economic growth. E.g. for the Sala-i-Martin et al. (2004) data these are initial GDP, primary education and the investment price. Furthermore, the coefficient to initial GDP is about twice as large compared to Sala-i-Martin et al. (2004) and hence the conditional β-convergence speed (see e.g. Barro, 1991) is about twice as high as found in Sala-i-Martin et al. (2004). Several dummy, political and other variables, most notably the East Asian dummy having highest posterior inclusion probability in Sala-i-Martin et al. (2004), are not significant. Qualitatively similar findings prevail also for the Fernandez et al. (2001) data. For the European regional data in particular human capital appears to be significantly related to growth. Our findings show the importance of appropriate conditioning, via inclusion of principal components of the large set of potential explanatory variables, in uncovering the variables important to explain economic growth. From a computational perspective it turns out that the specific choice of information criterion based weighting scheme has limited importance on the model averaging results. This holds true especially for the inclusion weight ranking of variables but to a large extent also for the estimated model average coefficients.

The paper is organized as follows. Section 2 describes the econometric methods used. Section 3 contains the empirical analysis and results and Section 4 briefly summarizes and concludes. Three appendices follow the main text. Appendix A briefly describes the European regional data, Appendix B collects some additional empirical results and Appendix C describes the computation of confidence intervals for frequentist model average coefficients.

2 Description of the Econometric Approach

Let $y \in \mathbb{R}^N$ denote the variable to be explained (in our application average per capita GDP growth for N countries respectively regions) and collect all explanatory variables in $X = [X_1 \ X_2] \in \mathbb{R}^{N \times k}$, with the core variables given in $X_1 \in \mathbb{R}^{k_1}$ and the auxiliary variables in $X_2 \in \mathbb{R}^{k_2}$ with $k = k_1 + k_2$. Without loss of generality we assume that all variables have zero mean, since in all growth regressions an intercept is typically included. As is well known, by the Frisch-Waugh theorem, the regression can equivalently be estimated with demeaned variables. The regression including all variables is given by

$$y = X_1\beta_1 + X_2\beta_2 + u. \tag{1}$$

The information for regression (1) contained in X_2 is equivalently summarized in the set of (orthogonal) principal components corresponding to X_2, i.e. in the set of transformed variables $\breve{X}_2 = X_2 O$, with $O \in \mathbb{R}^{k_2 \times k_2}$ computed from the eigenvalue decomposition of $\Sigma_{X_2} = X_2' X_2$ (due to the assumption of zero means):

$$\begin{aligned} \Sigma_{X_2} = X_2'X_2 &= O\Lambda O' = [O_1 \ O_2] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} O_1' \\ O_2' \end{bmatrix} \\ &= O_1\Lambda_1 O_1' + O_2 \Lambda_2 O_2', \end{aligned} \tag{2}$$

where $O'O = OO' = I_{k_2}$ and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_{k_2})$, $\lambda_i \geq \lambda_{i+1}$ for $i = 1, \ldots, k_2 - 1$. The partitioning into variables with subscripts 1 and 2 will become clear in the discussion below. From (2) the orthogonality of $\breve{X}_2$ is immediate, since $\breve{X}_2'\breve{X}_2 = \Lambda$.

Let us consider the case of multi-collinearity in X_2 first (which e.g. necessarily occurs when $k_2 > N$) and let us denote the rank of X_2 with r. Take $\Lambda_1 \in \mathbb{R}^{r \times r}$, hence $\Lambda_2 = 0$ and $X_2'X_2 = O_1\Lambda_1 O_1'$. The space spanned by the columns of $X_2 \in \mathbb{R}^{N \times k_2}$ coincides with the space spanned by the orthogonal regressors $\tilde{X}_2 = X_2 O_1 \in \mathbb{R}^{N \times r}$, i.e. with the space spanned by the r principal components. Thus, in this case regression (1) is equivalent to the regression

$$y = X_1\beta_1 + \tilde{X}_2\tilde{\beta}_2 + u \tag{3}$$

in the sense that both regressions lead to exactly the same fitted values and residuals. Furthermore, in case $[X_1 \ \tilde{X}_2]$ has full rank, regression (3) leads to unique coefficient estimates of β_1 and $\tilde{\beta}_2$. Since linear regression corresponds geometrically to a projection this is evident and of course well known.

Resorting to principal components, however, also has a clear interpretation and motivation in case of full rank of X_2 and hence of Σ_{X_2}. In such a situation replacing X_2 by the first r principal components $\tilde{X}_2$ leads to a regression where the set of regressors $\tilde{X}_2$ spans that r-dimensional subspace of the space spanned by the columns of X_2 such that the approximation error to the full space is minimized in a least squares sense. More formally the following holds true, resorting here to the population level.[2] Let $x_2 \in \mathbb{R}^{k_2}$ be a mean zero random vector with covariance matrix Σ_{X_2} (using here the same notation for both the sample and the population quantity for simplicity). Consider a decomposition of x_2 in a factor component and a noise component, i.e. a decomposition $x_2 = Lf + \nu$, with $f \in \mathbb{R}^r$, $L \in \mathbb{R}^{k_2 \times r}$ and $\nu \in \mathbb{R}^{k_2}$ (for a given value of r). If the decomposition is such that the factors f and the noise ν are uncorrelated, then $\Sigma_{X_2} = L\Sigma_f L' + \Sigma_\nu$. Principal components analysis performs an orthogonal decomposition of x_2 into Lf and ν such that the noise component is as small as possible, i.e. it minimizes $\mathbb{E}(\nu'\nu) = \mathrm{tr}(\Sigma_\nu)$. As is well known, the solution is given by $f = O_1' x_2$, $L = O_1$, with $O_1 \in \mathbb{R}^{k_2 \times r}$ and $\nu = O_2 O_2' x_2$, using the same notation for the spectral decomposition as above.

Therefore, including only r principal components $\tilde{X}_2$ instead of all regressors X_2 has a clear interpretation. The principal components augmented regression (PCAR) includes 'as much information as possible' (in least squares sense) with r linearly independent regressors contained in the space spanned by the columns of X_2. We write the PCAR as:

$$y = X_1\beta_1 + \tilde{X}_2\tilde{\beta}_2 + \tilde{u}, \tag{4}$$

neglecting in the notation the dependence upon the (chosen) number of principal components r but indicating with $\tilde{u}$ the difference of the residuals to the residuals of (3). Including only the information contained in the first r principal components of X_2 in the regression when the rank of X_2 is larger than r of course amounts to neglecting some information and hence leads to different, larger residuals. Thus, in comparison to the full regression (1), if it can be estimated, the PCAR regression will in general incur some bias in the estimates which has to be weighed against the benefits of a lower estimator variance. It is immediate that the choice of r is a key issue. The larger r, the more information is included but the fewer degrees of

[2] I.e. we now consider the k_2-dimensional random vector x_2 for which X_2, a sample of size N, is available.

freedoms are left (i.e. a lower bias but a higher variance) and multi-collinearity (since the λ_i are ordered decreasing in size) may become a problem.[3] Any choice concerning r is based on the eigenvalues λ_i, where 'large' eigenvalues are typically attributed to the factors and 'small' ones to the noise. The literature provides many choices in this respect and we have experimented with several thereof.[4] A classical, descriptive approach is given by the so-called variance proportion criterion (VPC),

$$r_{VPC(\alpha)} = \min_{j=1,\ldots,k_2} \left(j | \frac{\sum_{i=1}^{j} \lambda_i}{\sum_{i=1}^{k_2} \lambda_i} \geq 1 - \alpha \right), \tag{5}$$

with $\alpha \in [0, 1]$. Thus, $r_{VPC(\alpha)}$ is the smallest number of principal components such that a fraction $1-\alpha$ of the variance is explained. For our applications setting $\alpha = 0.2$, i.e. explaining 80% of the variance, leads to reasonable numbers of principal components included. In the context of growth regressions there is no underlying theoretical factor model explaining the second-moment structure of the auxiliary variables X_2 available, thus any choice has to a certain extent a heuristic character and has to trade off good approximation (necessary to capture the information contained in all explanatory variables for small bias) with a sufficiently small number of principal components (necessary for well-behaved regression analysis with low variance).

When computing the principal components from the regressors $X_2 \in \mathbb{R}^{N \times k_2}$ in our growth application, we split this set of variables in two groups. One group contains the quantitative or cardinal variables and the other includes the dummy or qualitative variables. We separate these two groups to take into account their different nature when computing principal components. For both groups the principal components are computed based on the correlation matrix of the variables. Computing the principal components based on the correlation matrix is especially important for the group of quantitative variables. These differ considerably in magnitude, due to their scaling which we keep unchanged for the Fernandez et al. (2001) and Sala-i-Martin et al. (2004) data to use exactly the same data as in these papers. Computing the principal components based on the covariance matrix leads in such a case to essentially fitting the 'large' variables, whereas the computation based on the correlation matrix corrects

[3] Note here that multi-collinearity cannot only appear within $\tilde{X}_2$ but in the joint regressor matrix $[X_1\ \tilde{X}_2]$. In any empirical application this can, however, be easily verified and if necessary remedied by removing some variables from the set of orthogonal regressors $\tilde{X}_2$ in order to have a well-behaved PCAR.

[4] In addition to the results reported in the paper the number of principal components has been determined using the testing approaches of Lawley and Maxwell (1963), Malinowski (1989), Faber and Kowalski (1997), Schott (2006) and Kritchman and Nadler (2008). In a variety of simulations, however, the VPC criterion and a simple eigenvalue test based on the correlation matrix (see below) have performed best.

for scaling differences and leads to a scale-free computation of the principal components. To be precise, a weighted principal components problem is solved in which the function to be minimized is given by $\mathbb{E}(\nu' Q \nu) = \text{tr}(Q \Sigma_\nu)$ with $Q = \text{diag}(\sigma^{-2}_{x_2,1}, \ldots, \sigma^{-2}_{x_2,k_2})$, neglecting here for simplicity the separation of the variables in X_2 in quantitative and dummy variables.[5] This leads to $f = O_1' Q^{1/2} x_2$, $L = Q^{-1/2} O_1$ and $\nu = Q^{-1/2} O_2 O_2' Q^{1/2} x_2$, i.e. the auxiliary regressors are given by $\tilde{X}_2 = X_2 Q^{1/2} O_1$.

For a chosen number of principal components, the PCAR (4) allows to estimate the conditional effects of the variables X_1 taking into account the relevant information contained in X_2 and summarized in $\tilde{X}_2$. As discussed in the introduction, one can also use (4) as a starting point to consider model averaging. By resorting to PCAR analysis, the number of regressions to be computed to estimate all sub-models is reduced from 2^k to 2^{k_1} if one computes all sub-models with respect to the core variables. The number of regressions can be reduced further by partitioning the set of variables $X_1 = [X_{11} \ X_{12}]$, with $X_{11} \in \mathbb{R}^{N \times k_{11}}$ included in each regression and $X_{12} \in \mathbb{R}^{N \times k_{12}}$, where $k_1 = k_{11} + k_{12}$, contains the variables in- or excluded in the sub-models estimated. This further reduces the number of regressions to be computed to $2^{k_{12}}$ and makes it even more likely that all sub-models can be estimated. As already mentioned in the introduction, the small number of models has the advantage, for both classical and Bayesian approaches that inference need not be based on estimation results obtained only on subsets of the model space with a focus on small models.[6]

We denote the sub-model regressions, based on the partitioning of (4) as

$$y = X_{11}\beta_{11}(j) + X_{12}(j)\beta_{12}(j) + \tilde{X}_2\tilde{\beta}_2(j) + \tilde{u}(j). \tag{6}$$

The sub-models $\mathcal{M}_j$ are indexed with $j = 1, \ldots, 2^{k_{12}}$, where $X_{12}(j)$ denotes subset j of X_{12}. The corresponding coefficient estimates are given by $\hat{\beta}(j) = [\hat{\beta}_{11}(j)' \ \hat{\beta}_{12}(j)' \ \tilde{\beta}_2(j)']' \in \mathbb{R}^{k_{11}+k_{12}+r}$. Here, with some imprecision in notation we include in $\hat{\beta}_{12}(j) \in \mathbb{R}^{k_{12}}$ zero entries corresponding to all variables not included in model $\mathcal{M}_j$, whereas in (4) the dimension of $\beta_{12}(j)$ equals the number of variables of X_{12} included. We are confident that this does

[5]Performing the spectral decomposition on a correlation matrix allows for another simple descriptive criterion concerning the number of principal components. By construction the trace of a correlation matrix equals its dimensions, i.e. is equal to k_2. Therefore, if all k_2 eigenvalues were equally large, they all would equal 1. This suggests to include as many principal components as there are eigenvalues larger than 1, i.e. to consider the eigenvalues larger than 1 as big and those smaller than 1 as small. The results correspond closely to those obtained with VPC_α with $\alpha = 0.2$.

[6]This statement has to be interpreted correctly: Inference is based on a different type of subset of the model space, since all information contained in X_2 is summarized in $\tilde{X}_2$ and taken into account. This conditional model space, after purging the effects of $\tilde{X}_2$, however, is then fully exhausted.

not lead to any confusion.[7] Furthermore, note already here that the regression including all explanatory variables, i.e. all variables in X_{12}, will be referred to as full model in the empirical application. Model average coefficients $\hat{\beta}^w$ are computed as weighted averages of the coefficient estimates of the sub-regressions, i.e.

$$\hat{\beta}^w = \sum_{j=1}^{2^{k_{12}}} w(j)\hat{\beta}(j), \tag{7}$$

with $0 \leq w(j) \leq 1$ and $\sum_{j=1}^{2^{k_{12}}} w(j) = 1$. We consider four different weighting schemes: equal weights, MMA weights as considered in Hansen (2007) and smoothed AIC (S-AIC) and smoothed BIC (S-BIC) weights considered by Buckland et al. (1997) and discussed in detail in Claeskens and Hjort (2008). Equal weighting assigns weights $w(j) = \frac{1}{2^{k_{12}}}$ to each of the models. By definition, this model averaging scheme does not allocate model weights according to any measure of quality of the individual models and thus serves more as a baseline averaging scheme. The other model averaging schemes base the model weights on different information criteria to give higher weights to models showing better performance in the 'metric' of the underlying information criterion. Hansen (2007), based on Li (1987), advocates the use of a Mallows criterion for model averaging that under certain assumptions results in optimal model averaging in terms of minimal squared error of the corresponding model average estimator amongst all model average estimators. The MMA model weights are obtained by solving a quadratic optimization problem. Denote with $\hat{U} = [\hat{u}(1), \ldots, \hat{u}(2^{k_{12}})] \in \mathbb{R}^{N \times 2^{k_{12}}}$ the collection of residual vectors of all models and with $M = [\dim(\mathcal{M}_1), \ldots, \dim(\mathcal{M}_{2^{k_{12}}})]' \in \mathbb{R}^{2^{k_{12}}}$ the dimensions of all models. The dimension of $\mathcal{M}_j$ is given by $k_{11}+r$ plus the number of variables of X_{12} included in $\mathcal{M}_j$. Further, denote with $\hat{\sigma}_F^2$ the estimated residual variance from the full model including all variables of X_{12}. Then, the MMA weight vector is obtained by solving the following quadratic optimization problem, where $w = [w(1), \ldots, w(2^{k_{12}})]' \in \mathbb{R}^{2^{k_{12}}}$ is the vector of weights corresponding to all models.

$$\min_w \left\{ w'\hat{U}'\hat{U}w + 2\hat{\sigma}_F^2 w'M \right\} \tag{8}$$

$$\text{subject to: } w \geq 0, \ \sum_{j=1}^{2^{k_{12}}} w(j) = 1.$$

[7] Note furthermore that we can, since $\tilde{X}_2$ are included in each regression, invoke the Frisch-Waugh theorem and entirely equivalently consider model averaging only for the regressions of y on X_{11} and the subsets of X_{12} by considering the residuals of the regressions of y, X_{11} and X_{12} on $\tilde{X}_2$. This equivalent interpretation highlights again that the inclusion of $\tilde{X}_2$ conditions on the 'relevant' information contained in X_2.

The remaining two averaging schemes base their weights on the information criteria AIC and BIC, defined here as $AIC(j) = N \ln \hat{\sigma}_j^2 + 2dim(\mathcal{M}_j)$ and $BIC(j) = N \ln \hat{\sigma}_j^2 + \ln N dim(\mathcal{M}_j)$, where $\hat{\sigma}_j^2$ is the estimated residual variance of $\mathcal{M}_j$. Based on these the corresponding model weights are computed as $w(j) = \exp\{-\frac{1}{2}AIC(j)\}/\sum_m \exp\{-\frac{1}{2}AIC(m)\}$ for S-AIC weights and as $w(j) = \exp\{-\frac{1}{2}BIC(j)\}/\sum_m \exp\{-\frac{1}{2}BIC(m)\}$ for S-BIC weights.

Each of the variables in X_{12} is included in exactly half of the models considered. The model average coefficient corresponding to each of the variables $X_{12,i},\ i = 1, \ldots, k_{12}$ can be written as

$$\begin{aligned} \hat{\beta}_{12,i}^w &= \sum_{j=1}^{2^{k_{12}}} w(j)\hat{\beta}_{12,i}(j) \\ &= \sum_{j:X_{12,i}\notin\mathcal{M}_j} w(j)0 + \sum_{j:X_{12,i}\in\mathcal{M}_j} w(j)\hat{\beta}_{12,i}(j). \end{aligned} \tag{9}$$

This shows the shrinkage character of model averaging. This is most clearly seen for equal weighting, for which the inclusion weight of variable i, i.e. $\sum_{j:X_{12,i}\in\mathcal{M}_j} w(j)$, is exactly 1/2 for all variables $X_{12,i}$. Hence for equal weighting the average coefficient is given by $\frac{1}{2^{k_{12}}}$ times the sum of all coefficient estimates over only $2^{k_{12}-1}$ (i.e. half of the) models. More generally, for any given weighting scheme the inclusion weight of variable i indicates the importance of this particular variable, in the 'metric' of the chosen weighting scheme. Thus, the inclusion weight is in a certain sense the classical alternative to Bayesian posterior inclusion probabilities. If the inclusion weight of a certain variable is high, this means that the 50% of the models in which this variable is included have a high explanatory power or good performance for e.g. with respect to AIC or BIC.

Given model average coefficients proper inference concerning them, e.g. to test for significance, is important. Correct statistical inference has to take into account that model averaging estimators are (random) mixtures of correlated estimators. Frequentist (or classical) inference taking these aspects into account has been developed in Hjort and Claeskens (2003) and is discussed at length in Claeskens and Hjort (2008, Chapter 7). A brief description of the computational aspects is contained in Appendix C and for further conceptual considerations we refer the reader to the cited original literature.

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 27 member states 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. The selection of core and auxiliary variables considered in this paper is based on the results obtained with the same data sets in Schneider and Wagner (2008). That paper follows a complementary approach to growth regressions in terms of obtaining point estimates of the coefficients to the relevant variables by resorting to adaptive LASSO estimation (see Zou, 2006). This estimation procedure performs at the same time consistent parameter estimation and model selection. We include (all respectively a subset of) the variables found important in that paper in our set X_1. For the Sala-i-Martin et al. (2004) and Fernandez et al. (2001) data sets the sets of variables found important in Schneider and Wagner (2008) are very similar to the sets of variables found important in the original papers based on Bayesian model averaging techniques, see the discussion in the respective subsections below. Thus, the sets X_1 include for these two data sets variables found to be important by studies using very different methods and thus constitute a potentially relevant starting point for applying the approach outlined in the previous section. Note that the choice of variables to be included in X_1 is here based on statistical analysis rather than being motivated by a particular economic theory model or question. By definition the results obtained with our approach depend upon the allocation of variables in the sets X_1 and X_2. Consequently, this is a key issue that deserves attention and our allocation based on the statistical analysis performed in Schneider and Wagner (2008) implies that the analysis performed and the results reported in this paper have to a certain extent illustrative character.

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.

As core variables (i.e. as regressors X_1) we consider 12 out of the 67 explanatory variables of the data set. In the list of variables to follow we include whether the estimated coefficients, the point estimates in Schneider and Wagner (2008, Table 2) and the posterior means of Sala-i-Martin et al. (2004, Table 4, p. 830), are positive or negative. The signs of the point estimates and the posterior means coincide for all variables. Furthermore we also include the rank with respect to posterior inclusion probability as given in Sala-i-Martin et al. (2004, Table 3, p. 828–829). Given that the largest part of the discussion in Sala-i-Martin et al. (2004) is for the results based on priors with mean model size 7 we compare our results throughout with these results of Sala-i-Martin et al. (2004). The variables are in alphabetical order of abbreviation: BUDDHA (fraction of population Buddhist in 1960, positive, 16), CONFUC (fraction of population Confucian in 1960, positive, 9), EAST (East Asian dummy, positive, 1), GDP (log per capita GDP in 1960, negative, 4), GVR61 (share of expenditure on government consumption of GDP in 1961, negative, 18), IPRICE (investment price, negative, 3), LAAM (Latin American dummy, negative, 11), MALFAL (index of malaria prevalence in 1966, negative, 7), P (primary school enrollment rate, positive, 2), REVCOUP (number of revolutions and coups, negative, 41), SAFRICA (sub-Saharan Africa dummy, negative, 10), TROPICAR (fraction of country's land in tropical area, negative, 5).[8]

We consider 3 out of the 12 variables to be included in all regressions, i.e. X_{11} = [EAST GDP P]. The East Asian dummy and the primary school enrollment rate are the two variables with the highest inclusion probabilities in Sala-i-Martin et al. (2004). Initial GDP is also found to be important and is by definition the central variable in the conditional β-convergence literature. Thus, in total only $2^9 = 512$ regressions are estimated for this data set. The total computation time is just a few minutes on a standard PC, showing that from a computational point of view also all $2^{12} = 8 \times 512 = 4096$ sub-model regressions could be estimated. The computationally most intensive part is actually the solution of the quadratic optimization problem to obtain the MMA weights. Out of the 55 variables in X_2, 11 are dummy variables (see Appendix B). Using the VPC criterion with 80%, 13 principal components are included

[8]In Schneider and Wagner (2008) 14 variables are selected by the adaptive LASSO algorithm. However, the coefficients for two of them are not significant, with standard errors computed as described in Zou (2006), in the final equation and have therefore been not included in X_1. These are GDE (average share of public expenditure on defense, positive, 45) and GEEREC (average share of public expenditure on education, positive, 48). As can be seen also their ranks with respect to posterior inclusion probability are rather high. Hence, these two variables are included in X_2.

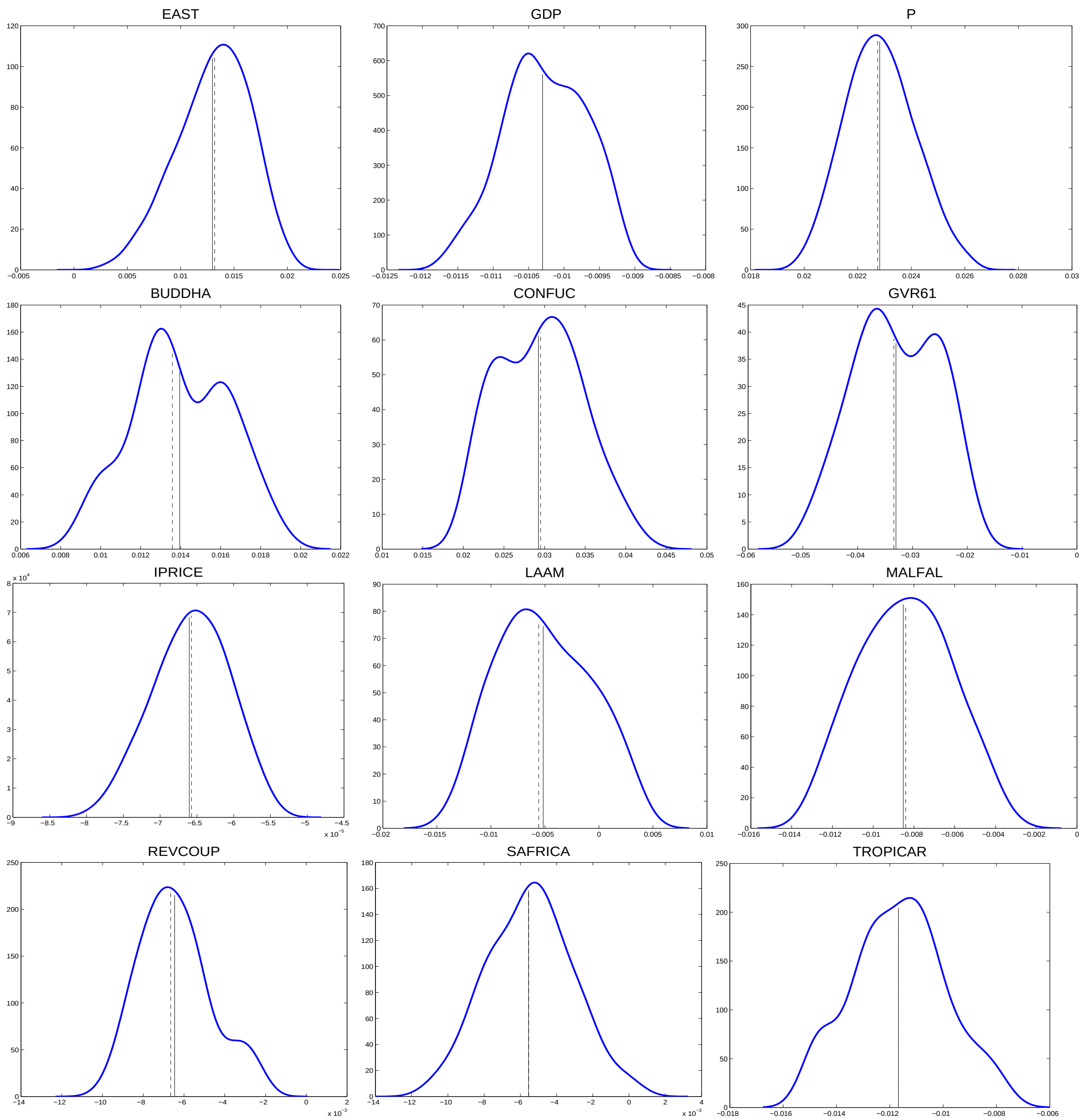

Figure 1: Empirical coefficient densities over all estimated models where the respective variables are included for the Sala-i-Martin et al. (2004) data set. The solid vertical lines display the means and the dashed vertical lines display the medians.

for the 44 quantitative variables and 6 principal components for the 11 dummy variables. Thus, in total 19 variables are included in $\tilde{X}_2$. Together, with 12 variables in X_1 the full regression includes 31 variables.[9] The smallest regression includes 22 variables, the three variables in X_{11} and the 19 principal components.

Figure 1 displays the empirical coefficient densities for the coefficients to the variables in X_1, ordered (alphabetically within groups) as first those in X_{11} and then those in X_{12}. All empirical coefficient densities displayed in this paper are based on Gaussian kernels with the bandwidths chosen according to Silverman's rule of thumb, see Silverman (1986) for details. For all coefficients to variables in X_{11} the densities are computed based on all $2^{k_{12}}$ available estimates, and for the coefficients to variables in X_{12} the densities are, of course, based on only the $2^{k_{12}-1}$ estimates in the models where the respective variables are included. Numerical information (mean, standard deviation and quantiles) concerning these empirical distributions is contained in table format in Table 9 in Appendix B. For some of the variables (BUDDHA, CONFUC, GVR61) bimodality occurs over the set of models estimated. It is important to note that these densities cannot be interpreted in a similar fashion as posterior densities in a Bayesian framework. The empirical densities merely visualize the variability of the estimated coefficients over all estimated models. These individual coefficients are then weighted with several weighting schemes to obtain model average coefficients. It is the unknown densities of the model average coefficients that are the classical counterparts to posterior densities.

The means displayed in Figure 1 correspond by construction to the model average estimates for the equal weights weighting scheme for the coefficients to the variables in X_{11} and are twice the means for the coefficients to the variables in X_{12} when the means are computed over all models, i.e. not conditional upon inclusion. Figure 2 displays the distribution of model weights over the model sizes for the four discussed model averaging schemes, see also Table 10 in Appendix B. This figure, as well as the similar ones for the other two data sets, displays for simplicity the model size ranging from 0 to $\dim(X_{12})$, i.e. 1 (for the intercept) plus the number of variables in X_{11} and the number of principal components included are not added on the horizontal axis. The 'real' model sizes are $1 + k_{11} + r = 23$ (i.e. the intercept, the variables X_{11} and the principal components) plus the numbers indicated in the figure.

[9] Using 31 explanatory variables, or 32 if the intercept (i.e. the demeaning) is counted as well, for 88 observations could be seen as too large a number. The number of variables can be reduced by explaining a smaller percentage than 80% of the variance of the auxiliary variables or by computing the principal components from all 55 variables together. We have experimented with both possibilities and have found that the results are very robust.

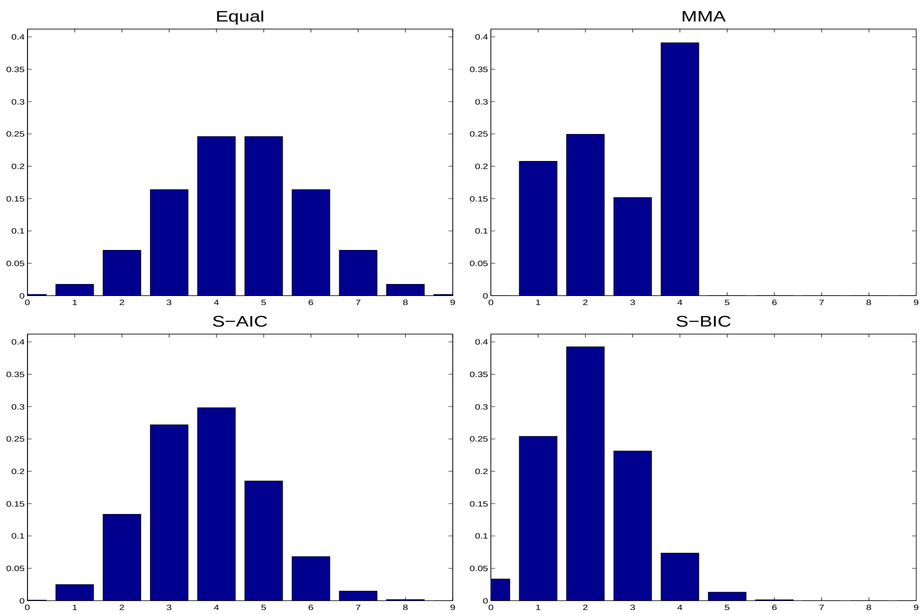

Figure 2: Distribution of model weights over model sizes for the Sala-i-Martin et al. (2004) data set. The weighting schemes displayed are: Equal (upper left graph), MMA (upper right graph), S-AIC (lower left graph) and S-BIC (lower right graph).

Alternatively, the numbers in the figure are the model sizes in the second regression, using again the Frisch-Waugh interpretation, after demeaning all variables and after conditioning on the information contained in $\tilde{X}_2$. For valid inference the real model size needs to be considered. By construction, the upper left graph simply displays the corresponding binomial weights but the other graphs are more informative. E.g. MMA averaging allocates all weights to model sizes ranging from 1 to 4 with 39% allocated to models with 4 variables included. The lower two graphs, corresponding to S-AIC and S-BIC model averaging, display that as expected S-BIC weighting allocates more weight on smaller models than S-AIC weighting. The model size with largest weight is 4 (30%) for S-AIC and 2 (39%) for S-BIC.

Table 1 displays the inclusion weights for the 9 variables in X_{12}. Two observations can be made. First, the numerical values of the inclusion weights differ across the three weighting schemes but the rankings almost perfectly coincide. The highest inclusion weight is obtained for IPRICE and is, depending upon weighting scheme, between 79% for MMA weights and 94% for S-AIC weights. The lowest inclusion weights are 0 for MMA weights for the variables GVR61, REVCOUP and SAFRICA. Second, the rankings differ from the ranking of these variables according to posterior inclusion probabilities in Sala-i-Martin et al. (2004). For

	BUDDHA	CONFUC	GVR61	IPRICE	LAAM	MALFAL
MMA	0.543 (3)	0.516 (4)	0.000 (7)	0.792 (1)	0.099 (6)	0.208 (5)
S-AIC	0.552 (3)	0.447 (4)	0.197 (8)	0.938 (1)	0.258 (6)	0.327 (5)
S-BIC	0.238 (3)	0.150 (5)	0.068 (8)	0.811 (1)	0.107 (6)	0.196 (4)
	REVCOUP	SAFRICA	TROPICAR			
MMA	0.000 (7)	0.000 (7)	0.569 (2)			
S-AIC	0.252 (7)	0.196 (9)	0.588 (2)			
S-BIC	0.080 (7)	0.063 (9)	0.389 (2)			

Table 1: Inclusion weights and ranks in brackets for the variables that are in- respectively excluded in model averaging for the three data dependent model averaging schemes for the Sala-i-Martin et al. (2004) data.

about half of the variables the same ranking as in Sala-i-Martin et al. (2004) (when ranking only these 9 variables) is found, namely for CONFUC (for MMA and S-AIC), GVR61 (for S-AIC and S-BIC), IPRICE, LAAM and TROPICAR. Note for completeness that the posterior inclusion probabilities of Sala-i-Martin et al. (2004) are for most variables relatively similar to the numbers reported in Table 1, in particular to the inclusions weights obtained with MMA or S-BIC weights, indicating that at the outset quite different approaches lead to rather similar results with respect to the importance of the inclusion of certain variables.

The next question addressed is now the contribution of the individual variables in terms of their coefficients. The estimation results for the full regression and the model average coefficients are displayed in Table 2. Significance for the full equation estimates is computed using OLS standard errors. For the model average coefficients inference is performed as developed in Claeskens and Hjort (2008) and as described briefly in the previous section and in Appendix C. Both in the full equation as well as for the model average coefficients only few variables are significant. In the full equation these are log per capita GDP in 1960 (GDP), the primary school enrollment rate (P) and the investment price (IPRICE). Furthermore two religion variables, the fraction of Buddhist in the population in 1960 (BUDDHA) and the fraction of Confucian in the population in 1960 (CONFUC), are significant at the 10% level in the full equation. When considering model average coefficients only GDP, P and IPRICE are significant, with the latter being significant only at the 10% level when using equal weights. Thus, only three key economic variables are found to be significantly related to economic growth when including the information contained in the auxiliary variables by including principal components. Note that the significance of variables is highly related to

	EAST	GDP	P	BUDDHA	CONFUC	GVR61
Full	0.006681	**-0.010399**	**0.021586**	*0.015432*	*0.036302*	-0.032472
Equal	*0.013071*	**-0.010273**	**0.022839**	0.006994	0.014720	-0.016476
MMA	0.012274	**-0.010362**	**0.023435**	0.009770	0.016824	0.000000
S-AIC	0.012248	**-0.010556**	**0.023508**	0.008714	0.013572	-0.004972
S-BIC	*0.014706*	**-0.010672**	**0.024268**	0.003625	0.004117	-0.001787
SDM04	0.017946	-0.005849	0.021374	0.002340	0.011212	-0.004594
SW08	0.013874	-0.001452	0.016097	0.012027	0.025531	-0.041727

	IPRICE	LAAM	MALFAL	REVCOUP	SAFRICA	TROPICAR
Full	**-0.000066**	-0.005742	-0.004173	-0.008964	-0.006988	-0.008287
Equal	*-0.000033*	-0.002533	-0.004214	-0.003212	-0.002737	-0.005824
MMA	**-0.000054**	-0.001078	-0.002521	-0.000000	-0.000000	-0.006852
S-AIC	**-0.000062**	-0.001925	-0.002445	-0.001682	-0.000863	-0.006821
S-BIC	**-0.000054**	-0.000873	-0.001826	-0.000481	-0.000246	-0.004844
SDM04	-0.000065	-0.001901	-0.003957	-0.000205	-0.002265	-0.008308
SW08	-0.000071	-0.002593	-0.001841	-0.002174	-0.002010	-0.005398

Table 2: Coefficient estimates for the Sala-i-Martin et al. (2004) data. *Full* displays the coefficient estimates corresponding to the full model; *Equal* the estimates corresponding to equal model weights; *MMA* the estimates using the weights as discussed in Hansen (2007); *S-AIC* the estimates computed with smoothed AIC weights and *S-BIC* the estimates computed with smoothed BIC weights. **Bold** typesetting indicates significance at the 5% significance level and *italic* numbers indicate significance at the 10% level, computed as discussed in Claeskens and Hjort (2008).
The rows labelled SDM04 display the unconditional posterior means of the coefficient estimates computed from Sala-i-Martin et al. (2004, Table 3, p. 828–829) and Sala-i-Martin et al. (2004, Table 4, p. 830) for mean prior model size 7. The rows labelled SW08 display the adaptive LASSO point estimates of Schneider and Wagner (2008, Table 2).

the inclusion weights, since amongst the variables in X_{12} the variable IPRICE has the highest inclusion weight.[10]

How do these findings relate to those in Sala-i-Martin et al. (2004)? Considering again model average coefficients, the three variables with significant coefficients are highly ranked in terms of posterior inclusion probability in Sala-i-Martin et al. (2004): GDP (4), P (2) and IPRICE (3). A key difference is that the variable with the highest inclusion probability, the East Asian dummy is not found to be significantly related to economic growth.[11] Also several other political, religious or health variables found to be important in Sala-i-Martin et al. (2004) are not significant in our results. Furthermore, the β-convergence speed implied by

[10] In supplementary material, available upon request, we provide for all three data sets also the model average estimates conditional upon inclusion.

[11] To be precise, the corresponding model average coefficient is significant at the 10% level for both equal and S-BIC weights.

our results is about twice as high as found by Sala-i-Martin et al. (2004), since our coefficient estimates for initial GDP are about twice as high in absolute value.

The differences in findings by construction originate in the differences in the approaches used. First, Sala-i-Martin et al. (2004) perform Bayesian model averaging over a different subset of the model space with small prior model sizes. Our PCAR approach controls for 80% of the variation in the auxiliary variables in *every* considered model. Thus, in each model the effect of the auxiliary variables is taken into account. This implies that those variables found to be insignificant in our results do not have high *additional* explanatory power relative to the information already taken into account in the auxiliary variables. At this point it is important to briefly discuss the selection of variables in X_1. Schneider and Wagner (2008) use the adaptive LASSO estimator, which is a specific penalized least squares estimator, to perform simultaneously model selection and parameter estimation. Thus, all variables in X_1 are variables that are found to be significantly related to GDP in a penalized regression framework. The variables in X_1 are typically also correlated with the principal components $\tilde{X}_2$, where e.g. the regression of EAST on the principal components leads to an R^2 of about 0.6. Using once again the Frisch-Waugh interpretation this implies that the residuals of a regression of EAST on the principal components are not significantly related to the residuals of the regression of GDP growth on the principal components in regressions where subsets of the other variables in X_1 are also included (or to be precise the residuals of regressions of these variables on the principal components are also included). Thus, with our approach exactly those variables are found to be significant that are related to GDP growth after conditioning on the auxiliary variables. These are the variables whose effect on growth is – in the sense discussed – well-distinguishable from the effects of other variables on growth. In slight abuse of commonly used notation, we can coin these variables as being 'robustly' related to economic growth.

Another important (computational) observation is that the choice of the weighting scheme (in particular MMA, S-AIC or S-BIC weights) has minor impact on both the significance as well as the numerical value of model average coefficients. The overall key finding of the application of our approach to the Sala-i-Martin et al. (2004) data is that we find *three key economic* variables significantly related to economic growth over the period 1960-1996, namely the logarithm of per capita GDP in 1960, the primary school enrollment rate and the investment price. These variables are significantly related to growth when controlling for the information contained in the auxiliary variables and are so in a model averaging framework where up to 9 additional variables, found to be important in Sala-i-Martin et al. (2004) and Schneider and Wagner (2008), are included as well.

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 Fernandez et al. (2001) 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. They add 16 further variables which are also available for these 72 countries, which gives a total of 41 explanatory variables. 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).

The choice of core variables is again based on Schneider and Wagner (2008). In the list of variables we include the sign of the point estimates obtained in Schneider and Wagner (2008, Table 4) and the ranks with respect to posterior inclusion probabilities from Fernandez et al. (2001, Table 1, p. 569).[12] In alphabetical order of abbreviation the list of variables in X_1 is: Confucius (share of population Confucian, positive, 2), EquipInv (equipment investment, positive, 4), EthnoLFrac (ethnolinguistic fractionalization, positive, 28), GDPsh560 (log of per capita GDP in 1960, negative, 1), HighEnroll (enrollment rates in higher education, negative, 34), LatAmerica (dummy for Latin America, negative, 13), LifeExp (life expectancy in 1960, positive, 3), Mining (fraction of GDP in mining, positive, 11), Muslim (share of population Muslim, positive, 6), NEquipInv (non-equipment investment, positive, 12), PrScEnroll (primary school enrollment in 1960, positive, 14), RuleofLaw (rule of law, positive, 7) and SubSahara (dummy for sub-Saharan Africa, negative, 5).

We include 4 out of the 13 variables in all regressions: Confucius, EquipInv, GDPsh560 and LifeExp. These are the four variables with the highest posterior inclusion probabilities in Fernandez et al. (2001). As before in total only $2^9 = 512$ regressions are estimated for this data set and again also all sub-model regressions could have been computed in terms of necessary computer time. Out of the 28 variables in X_2, 6 are dummies (see Appendix B) for which 4 principal components are included and 22 are quantitative variables for which 9

[12]Fernandez et al. (2001) do not report the posterior means of the coefficient estimates.

principal components are included, based again on the VPC criterion with 80%. Thus, for this data set 13 principal components are included and the real model sizes vary between 18 and 27 (counting also the intercept, i.e. the demeaning).

The discussion of results has the same structure as that in the previous subsection. Thus, we display in Figure 3 the empirical coefficient densities with the corresponding numerical information provided in Table 11 in Appendix B. In the figure first the 4 variables in X_{11} are displayed in alphabetical order and then the 9 variables in X_{12} are displayed in alphabetical order.

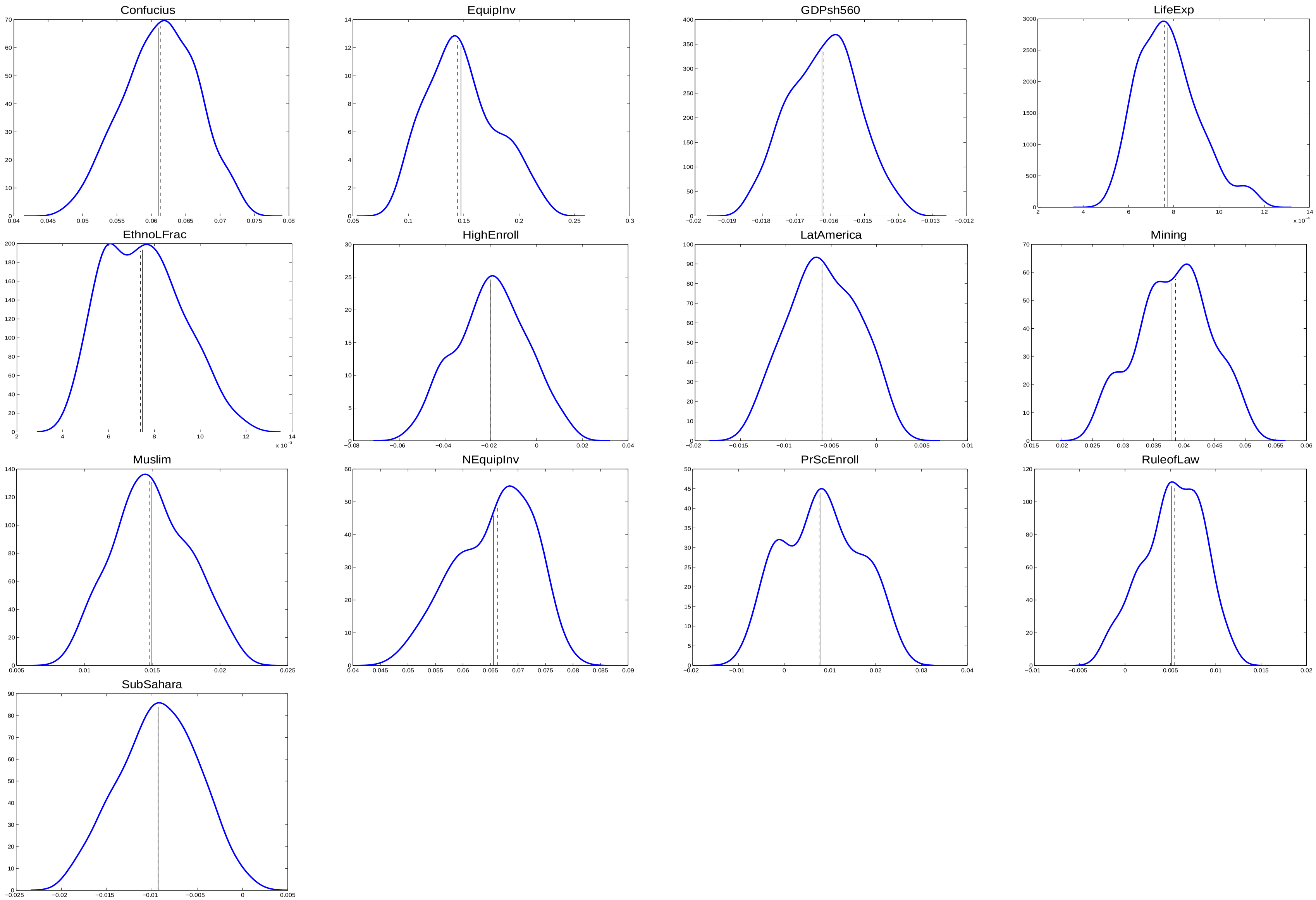

Figure 3: Empirical coefficient densities over all estimated models where the respective variables are included for the Fernandez et al. (2004) data set. For further explanations see caption to Table 1.

Figure 4 displays the model weights over model sizes, with the corresponding numerical information provided in Table 12 in Appendix B. With MMA weights all mass is allocated to models with sizes ranging from 2 to 4 and 6, with 54% of the weight given to models of size 4. S-AIC averaging allocates 33% to models of size 5 and S-BIC averaging allocates 46% to models of size 3.

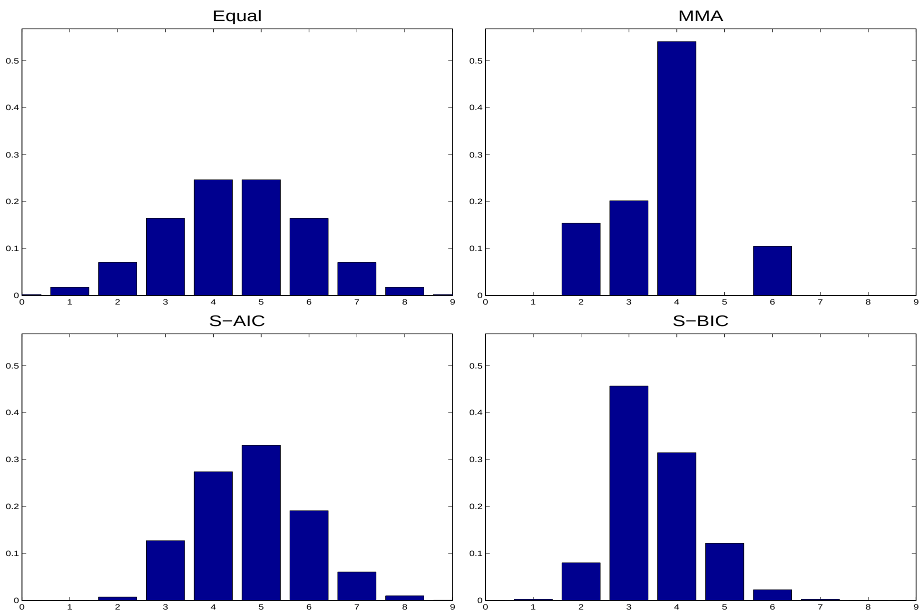

Figure 4: Distribution of model weights over model sizes for the Fernandez et al. (2001) data set. For further explanations see the caption of Table 2.

	EthnoLFrac	HighEnroll	LatAmerica	Mining	Muslim	NEquipInv
MMA	0.454 (4)	0.000 (9)	0.191 (7)	0.846 (1)	0.776 (2)	0.705 (3)
S-AIC	0.543 (4)	0.157 (9)	0.228 (7)	0.961 (3)	0.986 (1)	0.977 (2)
S-BIC	0.243 (4)	0.057 (9)	0.092 (7)	0.861 (3)	0.982 (1)	0.972 (2)
	PrScEnroll	RuleofLaw	SubSahara			
MMA	0.295 (6)	0.104 (8)	0.328 (5)			
S-AIC	0.371 (5)	0.214 (8)	0.356 (6)			
S-BIC	0.144 (5)	0.075 (8)	0.123 (6)			

Table 3: Inclusion weights and ranks in brackets for the variables that are in- respectively excluded in model averaging for the three data dependent model averaging schemes for the Fernandez et al. (2001) data.

For the inclusion weights of the variables in X_{12}, displayed in Table 3, again a large similarity in terms of ranking is observed for three considered weighting schemes. The ranks obtained with S-AIC weights and S-BIC weights coincide perfectly. Comparing these ranks with the posterior inclusion probability ranks reported in Fernandez et al. (2001) reveals some marked differences. The variable with highest posterior inclusion probability in Fernandez et al. (2001), SubSahara, is only ranked 5th (with MMA) or 6th with an inclusion weight of 36% for S-AIC weighting and even lower weight for the other weighting schemes. The third highest ranked variable in Fernandez et al. (2001), RuleofLaw, is ranked 8th. The three highest ranking variables (considering all three weighting schemes) are Mining, Muslim and NEquipinv. For this data set substantial differences occur between the inclusion weights computed here and the posterior inclusion probabilities of Fernandez et al. (2001).

	Confucius	EquipInv	GDPsh560	LifeExp	EthnoLFrac	HighEnroll
Full	**0.067374**	**0.120139**	**-0.017297**	**0.000781**	0.007952	-0.015508
Equal	**0.061269**	**0.148415**	**-0.016205**	**0.000776**	0.003744	-0.009845
MMA	**0.065054**	**0.131601**	**-0.016833**	**0.000845**	0.003657	0.000000
S-AIC	**0.067676**	**0.122169**	**-0.016956**	**0.000898**	0.003852	-0.000400
S-BIC	**0.066851**	**0.123147**	**-0.016823**	**0.000913**	0.001603	0.000158
SW08	0.056205	0.162509	-0.011495	0.000698	0.000373	-0.022536

	LatAmerica	Mining	Muslim	NEquipInv	PrScEnroll	RuleofLaw
Full	-0.002668	**0.038775**	**0.015167**	**0.048025**	0.011947	0.006376
Equal	-0.002913	0.019149	0.007497	0.032829	0.004046	0.002643
MMA	-0.002332	**0.033658**	**0.012843**	**0.046004**	0.005954	0.000840
S-AIC	-0.000893	**0.036223**	**0.015166**	**0.059591**	0.004180	0.000919
S-BIC	-0.000391	**0.031506**	**0.014977**	**0.063554**	0.001559	0.000285
SW08	-0.007402	0.019676	0.000222	0.000467	0.002728	0.005808

	SubSahara
Full	-0.008730
Equal	-0.004624
MMA	-0.003909
S-AIC	-0.002512
S-BIC	-0.000762
SW08	-0.018189

Table 4: Coefficient estimates for the Fernandez et al. (2001) data. The rows labelled SW08 display the adaptive LASSO point estimates of Schneider and Wagner (2008, Table 4). For further explanations see the caption to Table 2.

The estimation results displayed in Table 4 bear some qualitative resemblance to the findings of the previous subsection. First, there are only small differences across the different weighting schemes, with respect to both significance and numerical values of the coefficients. For this data set furthermore the findings from the full equation are quite similar to the model averaging results. Seven variables are found to be significantly related to economic growth. These are the four variables in X_{11}, namely Confucius, EquipInv, GDPsh560 and LifeExp. From the variables in X_{12} the three variables with the highest inclusion weights are significant, i.e. Mining, Muslim and NEquipInv. On the other hand some variables that have high inclusion probabilities in Fernandez et al. (2001) are not significant, most notably SubSahara ranked 5th in Fernandez et al. (2001). Thus, several key economic variables, equipment investment, initial GDP, the share of mining in GDP and non-equipment investment are found to be significantly related to economic growth.

Similarly to the Sala-i-Martin et al. (2004) data analyzed in the previous subsection our approach leads to insignificance of geographical dummies, here LatAmerica and SubSahara and EAST, LAAM and SAFRICA above. Again also several political or institutional variables are not found to be significant with our method, contrary to their alleged importance in Fernandez et al. (2001). Hence, as for the Sala-i-Martin et al. (2004) data also for the Fernandez et al. (2001) data our approach finds mainly key economic variables significantly related to growth, where for the latter data set additionally two religion variables, Confucius and Muslim, are significant.

3.3 European Regional Data

The third data set we analyze contains 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 very broad sense including also 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 large inter-country differences in the economic performance of the European regions we also include country dummies for the 19 out of the 27 countries that consist of more than just one region, see Table 7 in Appendix A. As for the other two data sets, the set of core variables is taken from Schneider and Wagner (2008). In alphabetical order these are given by: AccessRail (measure of accessibility by railroad, negative), ARL0 (activity rate of low educated in 1995, negative), Capital (dummy

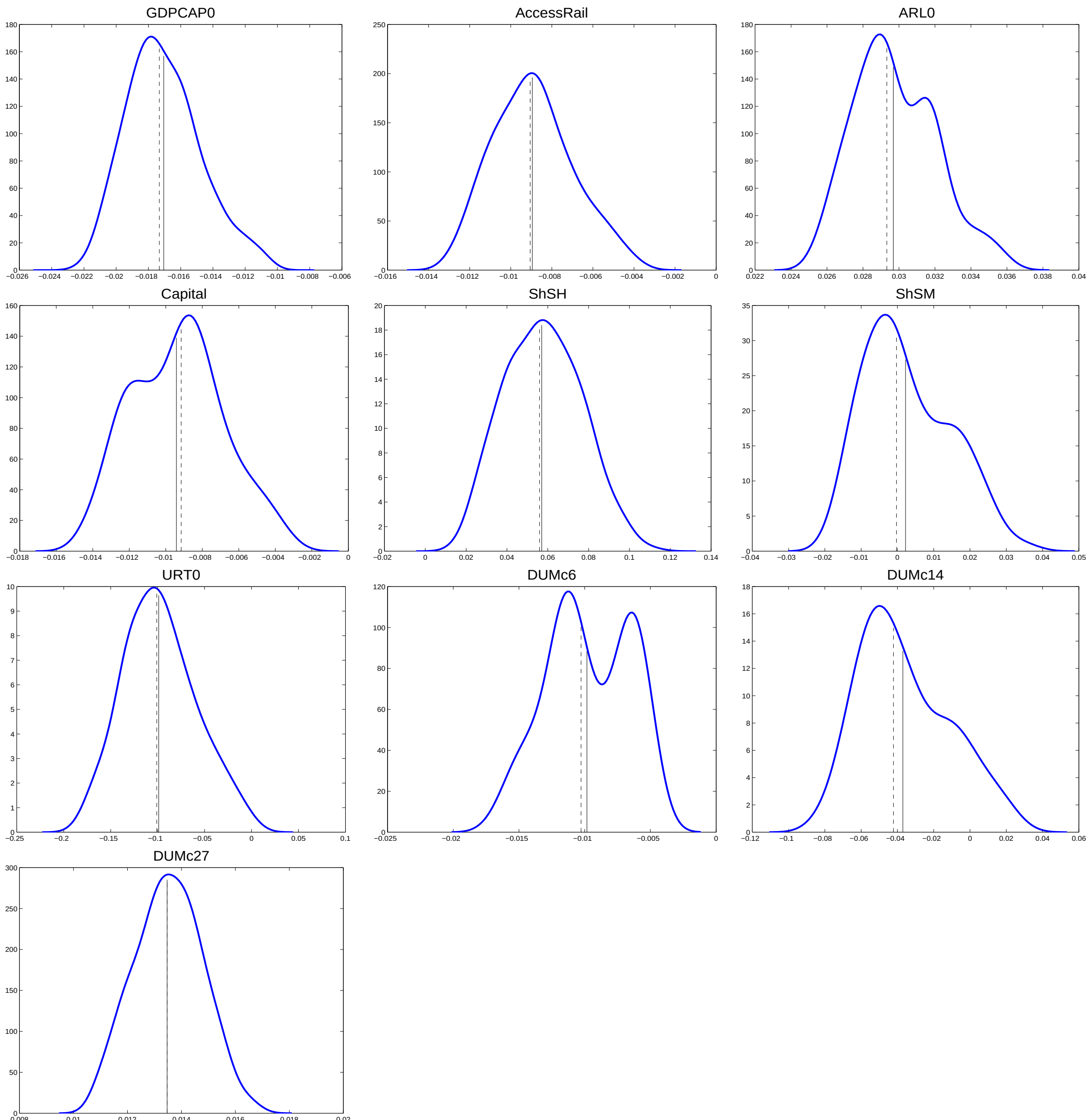

Figure 5: Empirical coefficient densities over all estimated models where the respective variables are included for the European regional data set. For further explanations see caption to Table 1.

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 included: DUMc6 (dummy for Germany, negative), DUMc14 (dummy for Ireland, positive) and DUMc27 (dummy for UK, negative). Positive respectively negative indicates the sign of the estimated coefficients in Schneider and Wagner (2008). For this data set we only include GDPCAP0 in all regressions and thus estimate again $2^9 = 512$ sub-model regressions. Out of the 57 variables in X_2, 23 are dummies (see Appendix B) for which 14 principal components are included and 34 are quantitative variables for which 10 principal components are included again based on the VPC criterion with 80%. Thus, for this data set for 255 regions 24 principal components are included and the real model sizes vary between 26 and 35 including the intercept.

Figure 5 displays the empirical coefficient densities and numerical information is provided in Table 13 in Appendix B. Bimodality occurs most markedly for the dummy variable for Germany (DUMc6) and to a lesser extent for the activity rate of low educated (ARL0) and the capital city dummy (Capital).

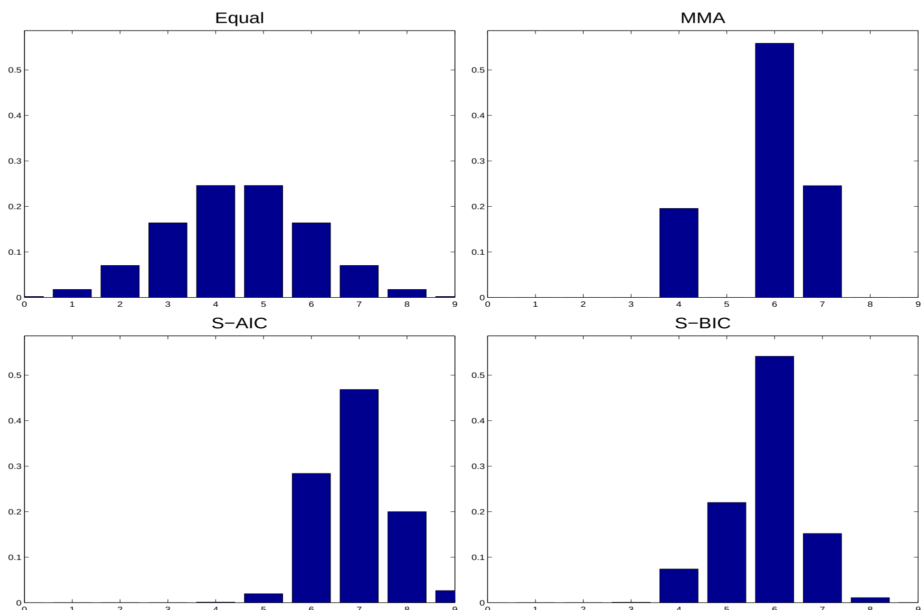

Figure 6: Distribution of model weights over model sizes for the European regional data set. For further explanations see the caption of Table 2.

With MMA weights only models of sizes 4, 6 and 7 have positive weights, as displayed in Figure 6 and Table 14 in Appendix B, with 56% allocated to models of size 6. S-AIC and S-BIC model averaging allocates all weights to models of sizes 4 and larger. The model size with highest weight share is 7 for S-AIC with 47% and 6 for S-BIC with 54%. As for the other two data sets the inclusion weights for the variables in X_{12} are in terms of ranking very similar for all three averaging schemes, see Table 5. The differences between MMA weighting on the one hand and S-AIC and S-BIC weighting on the other are a bit larger than the differences between S-AIC and S-BIC. The highest inclusion weights occur for the capital city dummy (Capital), the dummy for Germany (DUMc6) and the dummy for Ireland (DUMc14). Low inclusion weights are given to rail accessibility (AccessRail), the activity rate of low educated in 1995 (ARL0), the share of medium educated in the working age population (ShSM) and the total unemployment rate in 1995 (URT0). In addition to the three mentioned dummies also the share of high educated in the working age population (ShSH) has inclusion probability of above 80% with each of the weighting schemes.

	AccessRail	ARL0	Capital	ShSH	ShSM	URT0
MMA	0.491 (7)	0.000 (9)	0.851 (3)	0.804 (4)	0.288 (8)	0.712 (6)
S-AIC	0.520 (7)	0.191 (9)	1.000 (1)	0.980 (4)	0.334 (8)	0.925 (6)
S-BIC	0.151 (7)	0.054 (9)	1.000 (1)	0.844 (5)	0.094 (8)	0.870 (4)
	DUMc6	DUMc14	DUMc27			
MMA	0.971 (1)	0.933 (2)	0.804 (4)			
S-AIC	1.000 (1)	1.000 (1)	0.975 (5)			
S-BIC	1.000 (1)	1.000 (1)	0.791 (6)			

Table 5: Inclusion weights and ranks in brackets for the variables that are in- respectively excluded in model averaging for the three data dependent model averaging schemes for the European regional data.

The results obtained for the inclusion weights of the variables translate for this data set very clearly into significance of the estimated coefficients, with only small differences emerging between the different estimates. The coefficients corresponding to the two variables with the lowest inclusion probabilities, ARL0 and ShSM are both insignificant. The single exception being significance of ShSM when using MMA weights. Next to conditional convergence (due to the significant negative coefficient of initial GDP) at the speed of about 2% per year, important factors are being a capital city and – more policy relevant – a high share of high education in the labor force.

4 Summary and Conclusions

This paper offers two innovations for the empirical analysis of economic growth. First, it proposes the use of principal components augmented regressions (PCAR) for empirical growth analysis. This has several advantages, which include: First, PCAR analysis results in well-behaved regressions that include a large part of the information contained in the typically

	GDPCAP0	AccessRail	ARL0	Capital	ShSH	ShSM
Full	**-0.019748**	*-0.006189*	0.009345	**0.010931**	**0.064444**	0.015783
Equal	**-0.017046**	**-0.009724**	-0.035950	**0.013520**	**0.057723**	0.002732
MMA	**-0.019100**	**-0.008067**	-0.033232	**0.012289**	**0.065752**	**0.027556**
S-AIC	**-0.020046**	*-0.006138*	-0.001989	**0.011672**	**0.058074**	0.015685
S-BIC	**-0.019726**	*-0.006147*	-0.012906	**0.012535**	**0.056207**	0.017672
SW08	-0.014707	-0.001074	-0.004472	0.008078	0.0587	0.016212

	URT0	DUMc6	DUMc14	DUMc27
Full	**-0.100283**	**-0.012069**	**0.028863**	**-0.009459**
Equal	**-0.097582**	**-0.008916**	**0.029797**	**-0.009341**
MMA	**-0.115694**	**-0.010924**	**0.029372**	**-0.009878**
S-AIC	**-0.105123**	**-0.011198**	**0.029038**	**-0.008947**
S-BIC	**-0.115052**	**-0.010833**	**0.029430**	**-0.008426**
SW08	-0.005371	-0.007998	0.002764	-0.002237

Table 6: Coefficient estimates for the European regional data. The rows labelled SW08 display the adaptive LASSO point estimates of Schneider and Wagner (2008, Table 6). For further explanations see the caption to Table 2.

large sets of available variables. Second, this implies that the empirical analysis is based on regressions that suffer only from minor omitted variables bias and thus allows for more precise estimation of the conditional effects of the core variables on economic growth. Thus, well-defined estimates that take into account the theory open endedness of economic growth by conditioning on a large information set are obtained.

The second innovation of this paper is to consider frequentist model averaging instead of the usually employed Bayesian model averaging approaches. Inference for frequentist model average coefficients is based on recent advances in the statistics literature, in particular on Claeskens and Hjort (2008). We introduce amongst other quantities, the frequentist counterpart to the Bayesian posterior inclusion probability, which we term inclusion weight. In conjunction with PCAR model averaging becomes computationally very cheap, in either a Bayesian or a frequentist framework. The computations performed in this paper require only few minutes on standard PCs.

The proposed methodology is illustrated and 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 255 NUTS2 regions of the 27 European Union member states. The selection of core variables is for all three data sets based on the findings of Schneider and Wagner (2008), who use the adaptive LASSO estimator to study the determinants of economic growth. The findings are

very favorable and indicate that the proposed methodology is able to uncover economically relevant growth determinants with plausible coefficient estimates. For the example of the Sala-i-Martin et al. (2004) data, initial GDP, primary school enrollment and the investment price are found to be significant. The implied conditional convergence speed is about twice as high as found in Sala-i-Martin et al. (2004). Favorable findings are also obtained for the other two data sets. For the Fernandez et al. (2001) key economic variables found to be important are equipment investment, initial GDP, the share of mining in GDP and non-equipment investment. For both data sets several geographic dummy variables as well as other 'non-standard' economic variables (e.g. health and institutional variables), found to be important in the original studies, are not significant. This highlights that the improved estimation based on large sets of auxiliary conditioning variables results in plausible, economically relevant findings. For the European regional data set, for which only few key economic variables are available, the results highlight the importance of human capital for economic growth.
The findings in this paper forcefully illustrate that the proposed two methodological innovations, considered separately or together, are important additions to the toolkit of the empirical growth research community.

Acknowledgements

Financial support from the Jubiläumsfonds of the Oesterreichische Nationalbank (Project number: 11688) is gratefully acknowledged. The regional data set has been compiled and made available to us by Robert Stehrer from the Vienna Institute for International Economic Studies (WIIW). The comments of seminar participants at Universitat Pompeu Fabra, Barcelona, 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.

Brock, W.A. and Durlauf, S.N. (2001). Growth empirics and reality. *World Bank Economic Review* **15**, 229–272.

Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997). Model selection: An integral part of inference. *Biometrics* **53**, 603–618.

Claeskens, G. and Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.

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.

Faber, K. and Kowalski, B.R. (1997). Modification of Malinowski's F-test for abstract factor analysis applied to the quail root II data set. *Journal of Chemometrics* **11**, 53–72.

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

Hansen, B. (2007). Least squares model averaging. *Econometrica* **75**, 1175–1189.

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

Hjort, N.L. and Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association* **98**, 879–899.

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.

Kritchman, S. and Nadler, B. (2008). Determining the number of components in a factor model from limited noisy data. Working Paper, Weizmann Institute of Science.

Lawley, D.N. and Maxwell, A.E. (1963). *Factor Analysis as a Statistical Method*. London: Butterwoths.

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

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

Li, K.-C. (1987). Asymptotic optimality of C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Annals of Statistics* **15**, 958–975.

Malinowski, E.R. (1989) Statistical F-tests for abstract factor analysis and target testing. *Journal of Chemometrics* **3**, 49–60.

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.

Schneider, U. and Wagner, M. (2008). Catching growth determinants with the adaptive LASSO. Mimeo.

Schott, J.R. (2006). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *Journal of Multivariate Analysis* **97**, 827–843.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall CRC.

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

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. Most of the variables for which we report Eurostat as source 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.

Name	Description	Source
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
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

Continued on next page

Name	Description	Source
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
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: Additional Empirical Results

Sala-i-Martin et al. (2004) Data

Dummy variables, $0-1$ unless stated otherwise, contained in X_2: BRIT, COLONY, ECORG $(0, 1, ..., 5)$, EUROPE, LANDLOCK, NEWSTATE $(0, 1, 2)$, OIL, SCOUT, SOCIALIST, SPAIN, WARTORN

Variable	Mean	StdDev	5	10	25	50	75	90	95
EAST	0.013071	0.003405	0.006856	0.008415	0.010802	0.013349	0.015624	0.017295	0.018171
GDP	-0.010273	0.000575	-0.011287	-0.011034	-0.010668	-0.010276	-0.009802	-0.009481	-0.009374
P	0.022839	0.001283	0.020806	0.021162	0.021890	0.022770	0.023687	0.024555	0.025028
BUDDHA	0.013989	0.002399	0.009948	0.010515	0.012390	0.013618	0.015987	0.017266	0.017728
CONFUC	0.029440	0.005180	0.021583	0.022600	0.024895	0.029662	0.033451	0.036432	0.038295
GVR61	-0.032952	0.007627	-0.045244	-0.043221	-0.038594	-0.033052	-0.026328	-0.023067	-0.021360
IPRICE	-0.000066	0.000005	-0.000075	-0.000073	-0.000069	-0.000066	-0.000062	-0.000059	-0.000058
LAAM	-0.005066	0.004279	-0.011315	-0.010823	-0.008231	-0.005369	-0.001721	0.000917	0.002147
MALFAL	-0.008428	0.002248	-0.012161	-0.011454	-0.010130	-0.008289	-0.006770	-0.005322	-0.004599
REVCOUP	-0.006425	0.001747	-0.008918	-0.008525	-0.007719	-0.006598	-0.005430	-0.003737	-0.002967
SAFRICA	-0.005474	0.002326	-0.009282	-0.008474	-0.007171	-0.005454	-0.004017	-0.002487	-0.001566
TROPICAR	-0.011648	0.001738	-0.014704	-0.014373	-0.012877	-0.011609	-0.010536	-0.009301	-0.008502

Table 9: Mean, standard deviation and quantiles of the empirical coefficient distributions over all models where the respective variables are included for the Sala-i-Martin et al. (2004) data set.

	0	1	2	3	4	5	6	7	8	9
Equal	0.002	0.018	0.070	0.164	0.246	0.246	0.164	0.070	0.018	0.002
MMA	0.000	0.208	0.249	0.152	0.391	0.000	0.000	0.000	0.000	0.000
S-AIC	0.001	0.025	0.134	0.272	0.298	0.185	0.068	0.015	0.002	0.000
S-BIC	0.034	0.254	0.393	0.231	0.074	0.013	0.001	0.000	0.000	0.000

Table 10: Distribution of model weights over model sizes for the Sala-i-Martin et al. (2004) data for the four discussed weighting schemes.

Fernandez et al. (2001) Data

Dummy variables, $0-1$ unless stated otherwise, contained in X_2: BritCol $(0, 1, \ldots, 5)$, EcoOrg, FrenchCol, OutwarOr, SpanishCol, WarDummy

Variable	Mean	StdDev	5	10	25	50	75	90	95
Confucious	0.061269	0.005288	0.052543	0.053998	0.057679	0.061540	0.065271	0.067751	0.070608
EquipInv	0.148415	0.031306	0.100930	0.108573	0.125240	0.144671	0.169619	0.195328	0.205562
GDPsh560	-0.016205	0.000997	-0.017829	-0.017542	-0.016969	-0.016168	-0.015501	-0.014915	-0.014582
LifeExp	0.000776	0.000134	0.000586	0.000621	0.000669	0.000764	0.000852	0.000956	0.001025
EthnoLFrac	0.007488	0.001671	0.004990	0.005386	0.006021	0.007420	0.008627	0.009802	0.010384
HighEnroll	-0.019691	0.015222	-0.043586	-0.041503	-0.029917	-0.019753	-0.009870	-0.000387	0.006498
LatAmerica	-0.005826	0.003777	-0.012103	-0.010926	-0.008437	-0.005983	-0.002866	-0.000553	0.000136
Mining	0.038298	0.005875	0.027905	0.028967	0.034456	0.038741	0.042097	0.046711	0.048027
Muslim	0.014993	0.002767	0.010536	0.011204	0.013074	0.014802	0.017118	0.018776	0.019866
NEquipInv	0.065659	0.006905	0.053085	0.056004	0.060273	0.066609	0.071281	0.073975	0.074808
PrScEnroll	0.008091	0.008219	-0.004684	-0.002487	0.000673	0.007886	0.014713	0.019600	0.020849
RuleofLaw	0.005285	0.003215	-0.001024	0.000827	0.003198	0.005486	0.007921	0.009106	0.010260
SubSahara	-0.009249	0.004250	-0.016622	-0.014944	-0.012193	-0.009225	-0.006234	-0.003704	-0.002726

Table 11: Mean, standard deviation and quantiles of the empirical coefficient distributions over all models where the respective variables are included for the Fernandez et al. (2001) data set.

	0	1	2	3	4	5	6	7	8	9
Equal	0.002	0.018	0.070	0.164	0.246	0.246	0.164	0.070	0.018	0.002
MMA	0.000	0.000	0.154	0.202	0.540	0.000	0.104	0.000	0.000	0.000
S-AIC	0.000	0.000	0.007	0.127	0.274	0.330	0.191	0.061	0.010	0.001
S-BIC	0.000	0.003	0.080	0.456	0.315	0.122	0.023	0.002	0.000	0.000

Table 12: Distribution of model weights over model sizes for the Fernandez et al. (2001) data for the four discussed weighting schemes.

European Regional Data

Dummy variables, $0-1$ unless stated otherwise, contained in X_2: RegBorder, RegCoast, RegObj1, RegPent27, Seaports, Settl, TELF $(1, 2, \ldots, 6)$ and 16 country dummies for the countries consisting of more than one region, compare Table 7, with the exception of Germany (DUMc6), Ireland (DUMc14) and the UK (DUMc27). These three country dummies are included in X_{12} in the results discussed in detail in this paper.

Variable	Mean	StdDev	5	10	25	50	75	90	95
GDPCAP0	-0.017046	0.002289	-0.020494	-0.019834	-0.018679	-0.017266	-0.015663	-0.013885	-0.012638
AccessRail	-0.009724	0.003149	-0.015196	-0.014210	-0.011847	-0.010188	-0.006719	-0.005738	-0.005152
ARL0	-0.035950	0.024852	-0.069014	-0.065331	-0.054822	-0.041083	-0.017692	0.001011	0.010719
Capital	0.013520	0.001233	0.011404	0.011824	0.012646	0.013501	0.014391	0.015260	0.015413
ShSH	0.057723	0.018497	0.028123	0.032660	0.042673	0.056717	0.070823	0.081562	0.088858
ShSM	0.002732	0.011873	-0.013476	-0.011545	-0.006302	0.000250	0.012144	0.019983	0.024431
URT0	-0.097582	0.038242	-0.157835	-0.142333	-0.128205	-0.100769	-0.073134	-0.042282	-0.027830
DUMc6	-0.008916	0.001921	-0.011902	-0.011385	-0.010336	-0.008992	-0.007700	-0.006128	-0.005353
DUMc14	0.029797	0.002265	0.026481	0.026889	0.028159	0.029394	0.031490	0.032544	0.034152
DUMc27	-0.009341	0.002497	-0.013126	-0.012645	-0.011364	-0.009132	-0.007754	-0.006071	-0.004834

Table 13: Mean, standard deviation and quantiles of the empirical coefficient distributions over all models where the respective variables are included for the European regional data set.

	0	1	2	3	4	5	6	7	8	9
Equal	0.002	0.018	0.070	0.164	0.246	0.246	0.164	0.070	0.018	0.002
MMA	0.000	0.000	0.000	0.000	0.196	0.000	0.559	0.246	0.000	0.000
S-AIC	0.000	0.000	0.000	0.000	0.001	0.020	0.284	0.469	0.200	0.027
S-BIC	0.000	0.000	0.000	0.001	0.074	0.220	0.542	0.152	0.011	0.000

Table 14: Distribution of model weights over model sizes for the European regional data for the four discussed weighting schemes.

Appendix C: Inference for Model Average Coefficients

In order to describe how to perform inference as derived in Claeskens and Hjort (2008) some further quantities need to be defined first. Denote with $e_j \in \mathbb{R}^{k_{11}+r}$ a vector with 0 entries except for 1 at position j and with $\tilde{e}_j \in \mathbb{R}^{k_{12}}$ a vector with 0 entries except for 1 at position j. Next define $\tau_{0i}^2 \in \mathbb{R}_0^+$ and $\omega_i \in \mathbb{R}^{k_{12}}$ as

$$\tau_{0i}^2 = \begin{cases} e_i' \left(\begin{bmatrix} X_{11}' \\ \tilde{X}_2' \end{bmatrix} \begin{bmatrix} X_{11} & \tilde{X}_2 \end{bmatrix} \right)^{-1} e_i, & i = 1, \dots, k_{11} \\ 0, & i = k_{11}+1, \dots, k_{11}+k_{12} \\ e_{i-k_{12}}' \left(\begin{bmatrix} X_{11}' \\ \tilde{X}_2' \end{bmatrix} \begin{bmatrix} X_{11} & \tilde{X}_2 \end{bmatrix} \right)^{-1} e_{i-k_{12}}, & i = k_{11}+k_{12}+1, \dots, k_{11}+k_{12}+r \end{cases} \tag{10}$$

and

$$\omega_i = \begin{cases} X_{12}' \begin{bmatrix} X_{11} & \tilde{X}_2 \end{bmatrix} \left(\begin{bmatrix} X_{11}' \\ \tilde{X}_2' \end{bmatrix} \begin{bmatrix} X_{11} & \tilde{X}_2 \end{bmatrix} \right)^{-1} e_i, & i = 1, \dots, k_{11} \\ \tilde{e}_{i-k_{11}}, & i = k_{11}+1, \dots, k_{11}+k_{12} \\ X_{12}' \begin{bmatrix} X_{11} & \tilde{X}_2 \end{bmatrix} \left(\begin{bmatrix} X_{11}' \\ \tilde{X}_2' \end{bmatrix} \begin{bmatrix} X_{11} & \tilde{X}_2 \end{bmatrix} \right)^{-1} e_{i-k_{12}}, & i = k_{11}+k_{12}+1, \dots, k_{11}+k_{12}+r \end{cases} \tag{11}$$

Further, we need the block of the information matrix for the coefficient vector β in the full regression including $[X_{11}\ X_{12}\ \tilde{X}_2]$, given in partitioned format by

$$\mathcal{I} = \frac{1}{N\sigma^2} \begin{bmatrix} X_{11}'X_{11} & X_{11}'X_{12} & X_{11}'\tilde{X}_2 \\ X_{12}'X_{11} & X_{12}'X_{12} & X_{12}'\tilde{X}_2 \\ \tilde{X}_2'X_{11} & \tilde{X}_2'X_{12} & \tilde{X}_2'\tilde{X}_2 \end{bmatrix}. \tag{12}$$

In the computations the unknown quantity σ^2 is replaced by the estimate from the full model. The 2-2 block of $\mathcal{I}^{-1}$, i.e. the block at the position of $X_{12}'X_{12}$ in $\mathcal{I}$, is denoted by $(\mathcal{I}^{-1})_{(2,2)}$. Next denote the set of indices of variables of X_{12} included in $\mathcal{M}_j$ as S_j and its cardinality by $|S_j|$, i.e. $S_j = \left\{i_1, \ldots, i_{|S_j|}\right\} \subseteq \{1, \ldots, k_{12}\}$. Without loss of generality we index the model excluding all variables of X_{12} as $\mathcal{M}_1$. For all models except $\mathcal{M}_1$ define

$$\pi_j = \begin{bmatrix} \tilde{e}_{i_1}' \\ \vdots \\ \tilde{e}_{i_{|S_j|}}' \end{bmatrix} \in \mathbb{R}^{|S_j| \times k_{12}} \tag{13}$$

and $G(j) = \pi_j' \left(\pi_j \left((\mathcal{I}^{-1})_{(2,2)}\right)^{-1} \pi_j'\right)^{-1} \pi_j \left((\mathcal{I}^{-1})_{(2,2)}\right)^{-1} \in \mathbb{R}^{k_{12} \times k_{12}}$ for $j = 2, \ldots, 2^{k_{12}}$. For $j = 1$ we define $G(1) = 0^{k_{12} \times k_{12}}$.

Based on these quantities one can compute for each coordinate $i = 1, \ldots, k_{11} + k_{12} + r$ of the model average coefficient vector $\hat{\beta}^w$ a valid confidence interval for testing the hypothesis that $H_0 :\ \beta_i^w = \beta_{i,0}$. Specifically, it holds (compare Theorem 4.1 of Hjort and Claeskens, 2003) that

$$T_{n,i} = \sqrt{\frac{N}{\tau_{0i}^2 + \omega_i'(\mathcal{I}^{-1})_{(2,2)}\omega_i}} \left[\hat{\beta}_i^w - \beta_{i,0} - \omega_i' \left(\hat{\beta}_{12}^F - \sum_{j=1}^{2^{k_{12}}} w(j)G(j)\hat{\beta}_{12}^F\right)\right] \tag{14}$$

is asymptotically standard normally distributed under the null hypothesis for $i = 1, \ldots, k_{11} + k_{12} + r$, where $\hat{\beta}_{12}^F$ is the block of the estimated coefficients corresponding to X_{12} in the full model. Based on (14) one can calculate confidence intervals.

Authors: Martin Wagner, Jaroslava Hlouskova

Title: Growth Regressions, Principal Components and Frequentist Model Averaging

Reihe Ökonomie / Economics Series 236

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

ISSN: 1605-7996