GORDON S. LANG SCHOOL of BUSINESS AND ECONOMICS

Department of Economics and Finance

Gordon S. Lang School of Business and Economics

University of Guelph

Discussion Paper 2020-04

Factors that affect Students' performance in Science: An application using Gini-BMA methodology in PISA 2015 dataset

By:

Anastasia Dimiski
University of Guelph
adimiski@uoguelph.ca

UNIVERSITY of GUELPH

IMPROVE LIFE.

Department of Economics and Finance
Gordon S. Lang School of Business and Economics | University of Guelph
50 Stone Road East | Guelph, Ontario, Canada | N1G 2W1
www.uoguelph.ca/economics

Factors that affect Students' performance in Science: An application using Gini-BMA methodology in PISA 2015 dataset

ANASTASIA DIMISKI
University of Guelph

Abstract. Existing theoretical and empirical evidence on the determinants of students' performance is relatively short. Even more narrow is the literature that examines the impact of pre-primary education on students' academic performance. Relying on the first-of-its-kind of the 2015 wave data from the Programme of International Student Assessment (PISA), the present study thoroughly discusses the associations between Students' performance in Science and a set of variables that are classified into 14 categories, including attendance and non-attendance in pre-primary education. To implement this research question, Gini-BMA approach is employed, which accounts for theory uncertainty. It is found that, among the factors, attendance in pre-primary education (i.e. PC11) is a robust determinant of students' performance in science. However, this result is supported only under the Gini methodology.

Keywords: students' performance, pre-primary education, Gini regression coefficient, BMA methodology, PISA.

1 Introduction

Formal education is without any doubt one of the major concerns for policy makers, since it determines an individual's income and amplifies inequalities of economic and social opportunities. Without having access to affordable high-quality education, people that belong at the lowest quintile of the income distribution are doomed to remain idle and underdeveloped, with few chances for upward mobility. While the linkages between education and income inequality and education and gender inequality have been commented on and studied extensively, relatively little analysis is available on the inequality of education and more specifically on the factors that affect Students' performance.

In a recent study, Helal et. al. (2019) identify three classes of factors that lead to lower academic performance: the socio-demographic element, which includes all students from indigent socio-economic background and those with special entry requirements, the academic one which includes all students with limited access to the course resources and forum and the course assessment element, which refers to students with low level contributions to the course level activities or to students who study-off campus or part-time. According to Tinajero et. al (2012), Brazilian university students' academic achievement is significantly enhanced by cognitive style and learning strategies. Hanushek and Wößmann (2006) take into consideration institutional differences by splitting schools into differing and non-differing ability systems and examine their impact on Students' mean performance.

Among the factors that stand out for their influence on Students' performance is pre-primary education since it would ensure a smooth transition to primary education and would set the foundation for lifelong learning. However, this research question has not been adequately studied. The contribution of this paper is to understand the associations between Students' performance in Science and a set of variables that are classified into 14 categories, including attendance and non-attendance in pre-primary education. To test this, a large cross-national dataset, the 2015 round of the Programme for International Student Assessment (PISA) is applied, which includes measures of achievement in science, mathematics and reading among 15-year old students in 72 countries.

PISA focuses on young students' ability to use their knowledge and skills to meet real-life challenges rather than on their ability to master specific curricula. The data are carefully selected through a serious international effort, with the selected measures to be both valid and consistent among the countries. The participation of a large number of countries allows more general claims, while the considerable number of schools and students within each country permits to simplify complex relationships between individual and school characteristics. Students take different combinations of different tests, but also, they answer questionnaires that provide information about students' background, schools' system, learning environment and learning experiences.

Bayesian Model Averaging (BMA) methodology is applied on this dataset, which moves the focus of analysis from estimates obtained from a given model, to estimates that do not depend on a particular model specification but rather use information from all candidate models. Thus, inference is averaged over models, forming a weighted average of model specific estimates where the weights are given by the posterior model probabilities. This framework permits to consider a wider range of possible explanatory variables and to end up with those that can effectively explain the relationship. To estimate the coefficients, Gini regression methodology is incorporated into the BMA. The Gini methodology is a rank-based methodology that takes into consideration both the variate values and their ranks and it is based on the Gini Mean Difference (GMD) as a measure of dispersion. Later, the results are compared to the OLS ones, which are taken by default when applying BMA methodology.

The contribution of this paper to the current literature is fourfold. First, it is an important contribution to the narrow literature that focuses on the factors that affect Students' performance in science. For this purpose, 43 variables, classified into 14 categories, are taken simultaneously into consideration. Most importantly, it attempts to shed a light on the following question: Does pre-primary education consist a crucial factor for Students' performance in science, and if yes, under what conditions? Second, this is among the very first studies that exploit OECD's PISA 2015 dataset. Third, this paper is one of the few studies to use Bayesian Model Averaging framework to answer the aforementioned research questions. Finally, it is the first time that Gini regression methodology is incorporated into the BMA one, to calculate the variables' coefficients.

The remainder of this paper is structured as follows. The second section is the literature review for the pre-primary education. The third section presents the literature review for BMA methodology. The forth section describes the BMA methodology with particular emphasis on prior model and parameter structures. The fifth section describes the theory for Gini regression coefficient. The sixth section describes the PISA data. The seventh section discusses the empirical results. The final section concludes.

2 Literature Review for pre-primary education

A growing literature is increasingly acknowledging the importance of early childhood interventions as an indispensable tool in nations building, as it has been argued that early interventions determine educational and labour market outcomes later in life (Cunha, Heckman, Lochner and Masterov,2006). As early childhood is considered a susceptible period for brain development and language acquisition (Heckman, Krueger and Friedman (2002); Knudsen, Heckman, Cameron and Shonkoff (2006)), pre-primary education assures a smooth transition to primary education and establishes the basis for later learning. A study by Carniero and Heckman (2003) points out that investments in human capital have dynamic complementarities and that skills obtained early in the child's lifetime expedite the development of additional future skills. So, early learning makes subsequent learning easier and generates important benefits in terms of medium and long-term schooling and socio-economic outcomes.

Early exposure to pre-primary schooling engenders supportive environment for the new intakes to easily adjust to formal school and develop essential social skills that lead to peer acceptance and academic achievement. (Myers 1992; Knight and Hughes 1995). Evidence abounds in the literature of the direct link between pre-school experience and academic performance. Entwisle and Alexander (1993) relate later school achievements to the children's academic skills obtained at school entry, while Berlinski et. al. (2009) links pre-primary school education to short-term gains in test scores and behavioral outcomes (e.g. attention, class participation, effort, and discipline). However, as is indicated by Behrman and Birdsall (1983), focusing exclusively on the quantity of pre-schooling might lead to misleading results because the variation in quality is substantial too. Using five different structural quality indicators, Bauchmüller et. al. (2014) find persistent, although modest, positive relationships between high quality early childhood care and children's test outcomes at the end of the primary school's 9th grade. In contrast, Chetty et.al. (2011) argues that high quality has a positive impact in cognitive development but is not lasting, since it fades out after few years. Goodman and Sianasi (2005) find that early education is related to improvements in cognitive skills at age 7, but the impact is short-lived since it remains important throughout the schooling years up to age 16. Similarly, using data from the Early Childhood Longitudinal Study, Manguson et. al. (2007) show that pre-school enrolment in the United States is associated with higher reading and mathematics skills at the time of entry into the first grade, but these effects dissipate for most children by the end of the first grade.

There are several reasons that justify the diminishing trend of/in gains from early childhood interventions. Esping-Andersen et. al. (2012) and Reynolds (1993,2000) state that children at risk due to family's low-income, poverty and other related factors cannot secure a continuous development if there is no a coherent, continual and adequate support provided by government funded preschool and primary grade intervention programs. Specifically, Zigler and Berman (1983) mention that a one-year intervention cannot "inoculate a child against continuing disadvantage" (p.898). Barnett (2011) mentions that interventions are not compelling when graduates from the early educational intervention programs attend public schools with limited efficiency. Further, Schulman et. al. (1999) and Barnett et.al (2004) acknowledge that although most of the states, across the United States, have established prekindergarten curriculum standards, they differ in terms of quality, accessibility, and availability of resources. Most importantly, few of them have established mechanisms to implement these comprehensive standards/prekindergarten initiatives.

Along with the early childhood interventions, many studies have found that home conditions are another crucial determinant of child's educational achievement (Bjorklund and Salvanes,2011). Both Velez et.al (1993) and Wößmann (2005) agree that, apart from preschool attendance, parental involvement and family features are key components in students' performance. A child's development begins within the family and depends on the parents' educational and cultural levels (Wößmann, 2005). Waldfogel and Washbrook (2011, and press b) support that parents that are educated and receive high income, spend more time to prepare their childrens' reading skills. In contrast, parents with lower income and less education have more possibilities to engage in harsh and incompatible parenting teaching behaviours that may negatively affect child's progress. Becker (1981, 1985) and Becker and Tomes (1986) embrace the theory of family to provide a reasonable justification for the failure of preschool education. Many authors correlate family's income with the quality of pre-school education too (Bainbridge et al., 2005; Magnuson and Waldfogel, 2005; Meyers et al.,2004). Low-income families are less likely to enrol their children to pre-school care, and if they do, they are most likely to be characterized by low-quality. In contrast, children from prosperous families are more likely to be registered in high-quality pre-schools. Attending systematically poorer quality pre-schools is an additional reason why gains from pre-school may eventually fade (Esping-Andersen et. al.,2012).

Expanding pre-primary education is an effective instrument to improve school progression and raise average achievement for less advantaged children. Extensive research has been conducted both on the short-run and long-run effects (see among others Barnett (1992), Barnett (1995), Danziger and Waldfogel (2000), Currie (2001), Blau and Currie (2006), Ludwig and Miller (2007)). Dumas and Lefranc (2010) find that extending pre-school enrolment in France is beneficial in terms of schooling outcomes, including test scores, for children from disadvantaged households. Heckman et. al. (2013) evaluate the results of the early childhood education Perry Preschool program that targeted to children from economically disadvantaged families. Outcomes reveal that children who participated in this program tended to create improvements in personality skills and enhance academic motivation. In particular, there is a boost in the long-term achievement test scores, with the effect being stronger for girls than for boys. Research suggests that disadvantaged children take the greatest advantage if, these special programs are of high quality (Gormley et al. (2005), Heckman and Lochner (2000), Neuman, Kamerman, Waldfogel, and Brooks-Gunn (2003), Reynolds et al. (2011), Waldfogel (2006)). Although there is ample empirical evidence that early childhood intervention programs have significant positive effects on the results of children from disadvantaged or minority background, it is not clear whether such pre-school programs influence the outcomes of children in the population as a whole. As typical preschool or prekindergarten programs vary in the quality of learning environments they provide and in the availability of financial resources, little is known about whether universal intervention can promote children's cognitive and academic outcomes (Gilliam and Zigler (2001)).

Many recent papers document the effects of universal preschool enrolment on the education of children in the entire population in a variety of other counties. Estimates obtained for developing counties testify positive and in some cases long-lasting effects of preschool attendance. Exploiting the information given by the Uruguayan Household Survey, Berlinski et. al. (2008) notice that attendance in pre-primary education reduces the probability for grade retention, grade failure and early drop-out during the primary and secondary schooling years. Aguilar and Tansini (2012) recognize that early exposure to pre-primary education has a positive effect on children's performance in the first year at public schools in Montevideo, Uruguay, and this effect remains positive but weakens after six years. Berlinski et.al. (2009) study the effects of Argentina's expansion of universal pre-primary schooling and find that pre-primary education positively affects third grade standardized Spanish and Mathematics test scores as well as students' behavioral skills. Taiwo and Tyolo (2002) notice that first grade Botswana students with pre-school experience achieve higher scores in English language, mathematics and science compared to students without such an experience. Using data for Thailand obtained from the Programme of International Student Assessment (PISA) for the years 2009 and 2012, Pholphirul (2017) reveals that pre-schooling attributes positively on cognitive skills in reading, mathematics and science with the mother's education attainment being a decisive factor on child's enrollment to preschool. According to this information, early exposure to pre-primary schooling appears as a successful and cost-effective policy to prevent late entry, early drop-out rates and early grade failure in poor countries, where large share of young population is excluded from compulsory education already at an early age (UNESCO, 2005).

There is considerable evidence for the impact of universal early childhood schooling in developed countries too. Using Census data, Cascio (2009) examines the long-run results of an expansion in universal kindergarten in the late 60s and early 70s across the United States. She reports no effect on the labour market outcomes and regarding the educational ones, the only positive influence is the reduction in grade retention. Goux and Maurin (2008) apply a difference-in-difference approach and find that one additional year in pre-elementary school in French has no important effect on children's subsequent educational skills. Baker et. al. (2005) show that the establishment of full-time and highly- subsidized kindergartens in the Canadian province of Quebec in the late 1990s, corresponds to an increase in the labour supply by married women and a decline in children's outcomes. Similarly, Dickson (2012) displays that the extension of free early education in the UK to all three -year- olds does not have any impact on reading, writing and mathematics when children reach the age seven. Only for deprived Local Education Authorities the results turn to be positive. In contrast Gormley and Gayer (2005) find that Oklahoma's universal pre-school program contributes positively to cognitive scores. In Japan, the expansion of both kindergartens and nursery schools is associated with higher achievement rates both in high school and college (Akabayashi and Tanaka (2013)).

3 Literature Review for Bayesian Model Averaging (BMA) approach

Classical Statistical Analysis disregards the theory and specification uncertainty, which jointly refer to as model uncertainty. As indicated by Leamer (1983), whimsical/arbitrary decisions about choice of functional forms and control variables leads to fragile inferences based on economic data. Bayesian Model Averaging (BMA) has successfully addressed model uncertainty in the model selection process, providing a comprehensible mechanism to embody ambiguity into conclusions about parameters. To construct estimates, it does not condition on a specific set of theories and covariates, but rather extracts information from a universe of candidate models. The result is a weighted average of model specific estimates, where posterior model probabilities are employed as weights.

The initial approach to deal with model averaging dates back to Roberts (1965)[1], who proposed a marginal distribution for outcomes of any unobserved sample, the so-called "predictive distribution". This distribution is defined as the weighted averaged of posterior probabilities of two models. Building up to this idea, Leamer (1978,1983) presented Extreme Bound analysis and set the fundamentals for the BMA methodology[2]. This technique was further studied by Raftery (1988, 1993), George and McCulloch (1993), Madigan and Raftery (1994), Drapper (1995), Kass and

[1] Regarding model selection, an initial approach is given by Efroymson (1960) who introduced the stepwise regression analysis.

[2] Levine and Renelt (1992) apply Extreme Bound Analysis to cross-country data and conclude that very few or none of the regressors robustly affect growth. To overcome possible difficulties arising from the implementation of this methodology, alternative solutions have been suggested (Leamer (1985), Granger and Uhlig (1990), Sala-i-Martin (1997)).

Raftery (1995), Kass and Wasserman (1995), Raftery, Madigan and Hoeting (1997), Hoeting et all (1999) among others.

BMA has proven a valuable tool in empirical settings with alternative theories, transmission mechanisms, a massive number of covariates and a limited number of observations. Empirical growth theory is a characteristic example. Theory uncertainty appears extensively in growth regressions since different theories, which use a specific subset of regressors, cannot exclude other compatible and interrelated theories from also being suitable. Brock and Durlauf (2001) refer to this by theory open-endedness, implying that existing growth theories needs to be considered when the effect of a specific growth theory on growth is to be analyzed. Apart from the aforementioned, Fernandez, Ley and Steel (2001a), Doppelhofer, Miller and Sala-i-Martin (2004), Durlauf, Kourtellos and Tan (2008), Durlauf, Kourtellos and Tan (2012), Massanjala and Papageorgiou (2008), Eicher, Papageorgiou and Raftery (2011), Ley and Steel (2007) are among those who endorse the use of this methodology to generate robust growth determinants.

Economic forecasting is another field which is affected by model uncertainty, since forecasts often depend on the model selected. Insightful contributions of forecast uncertainty can be found in Garratt, Lee, Pesaran and Shin (2003), Min and Zellner (1993), Raftery, Madigan and Hoeting (1997), Fernandez, Ley and Steel (2001 a, b), Ley and Steel (2009). In the context of forecasting inflation, Koop and Korobolis (2012) apply dynamic model averaging and selection strategies in state-space models to allow for both the coefficients in each model and the forecasting model to evolve over time. An alternative approach to predict inflation is used in Eklund and Karlsson (2007) who replace the standard marginal likelihood with the so-called forecast weights.

The rapid utilization of BMA in a variety of economic applications include examples from policy evaluation (Brock, Durlauf and West (2003), Sirimaneetham and Temple (2006), Kourtellos, Stengos and Tan (2013)), inequality (Kourtellos and Tsangarides (2017)), monetary policy (Levin and Williams (2003)), environmental economics (Begun and Eicher (2008)), returns to education (Tobias and Li (2004)) and intergenerational mobility (Kourtellos, Marr and Tan (2015)).

4 Bayesian Model Averaging (BMA) methodology

4.1 Basic BMA methodology

BMA provides a probabilistic framework to simultaneously deal with model and parameter uncertainty. To describe the relationships between all the unknown parameters and the data, a joint probability distribution is needed. To construct estimates, instead of conditioning on a single model, a model space $M = \{M_1, \ldots, M_k\}$ is taken into consideration, whose elements cover all the possible regressors suggested by the literature. For multiple model setups, it proceeds by assigning prior probability distributions to each model and to the parameters of each model. Combining those priors with the distribution for the data and conditioning on the data, results in the posterior distribution of model uncertainty, which allows for model selection and inferences[3].

Considering the case of normal linear regression models, model uncertainty occurs from the selection of the "best'' model, or alternatively, from the selection of the explanatory variables to include in the right-hand side:

$$Y = \beta_0 + \sum_{j=1}^{q} \left(\beta_j X_j + \varepsilon\right) = XB + \varepsilon\ , \quad \varepsilon \sim N(0, \sigma^2 I_n) \tag{1}$$

where Y and ε are nx1 vectors of the dependent variable and the error term respectively, X is a nxq matrix of candidate regressors that may or may not be included in the model, B is an nxq matrix with the parameters to be estimated and n is the total number of observations. If some of the elements of the parameter vector, $\beta = (\beta_1, \beta_2, \ldots \beta_q)$ equal zero, there are 2^q candidate models in total to be estimated, indexed by M_k for $k = 1, \ldots, 2^q$. Each of these models offers to explain the data and represents a distinct subset of the candidate regressors. For instance, model M_k takes the form:

$$Y = \sum_{j=1}^{qk} \beta_j^{(k)} X_j^{(k)} + \varepsilon \tag{2}$$

where Y is the dependent variable, ε is the normal error term, $X_1^{(k)} \ldots, X_{qk}^{(k)}$ is a subset of $X_1, .. X_q$ and $\beta = (\beta_1^{(k)}, \ldots, \beta_{qk}^{(k)})$ is a vector of regression coefficients to be estimated. The vector $\theta_k = (\beta_0, \beta^{(k)}, \sigma)$ summarizes the parameters for the given model M_k. By attaching a prior probability $p(M_k)$ to each model and a prior probability distribution $p(\theta_k \mid M_k)$ to the parameters of each model, a joint distribution over the models, the data and the parameters can be expressed as follows:

$$p(D, \theta_k, M_k) = p(D|\theta_k, M_k) p(\theta_k|M_k) p(M_k) \tag{3}$$

where $p(D \mid \theta_k, M_k)$ represents the likelihood function of model Mk which embodies all the information about the vector θ_k that is given by the data, D[4].

---

[3] For an excellent and detailed explanation of Bayesian model averaging see Raftery, Madigan, Hoeting (1997), Hoeting, Madigan, Raftery and Volinsky (1999), Sala-i-Martin (1997), Clyde and George (2004), Chipman, George, McCulloch (2001).

[4] According to Chipman et.al. (2001) and Clyde and George (2004), this prior information incorporates all these different models into one hierarchical mixture model (or, full model otherwise). The procedure to realize the data consists of three phases: first, the prior model distributions $p(M_1), \ldots p(M_k)$ are used to generate the model M_k. After deciding for the model, the parameter vector θ_k is generated using the prior probability distribution $p(\theta_k \mid Mk)$. In the last step, the likelihood function $p(D \mid \theta_k, M_k)$ of the selected model M_k is employed to generate the data D.

Establishing this prior formulation, the model uncertainty problem becomes that of finding the model M_k that was generated by the prior model probabilities $p(M_1),\ldots,p(M_k)$ and actually generated the data D. Conditioning on the observed data and applying the Bayes' theorem, the probability that M_k is the true model is given by the posterior model probability:

$$p(M_k|D) = \frac{p(D|M_k)\, p(M_k)}{\sum_k p(D|M_k) p(M_k)} \tag{4}$$

where p (D | M_k) is the marginal or integrated likelihood of model M_k, given, after applying the law of total probability, by[5]:

$$p(D|M_k) = \int p(D|\ \theta_k, M_k)\ p(\theta_k|M_k) d\theta k \tag{5}$$

The posterior model probability summarizes all the information contained in the data and presents a complete and consistent outline of post-data uncertainty. Thus, it can be treated as a measure of support for model Mk and it can be used as a model weight in BMA.

The posterior distribution of a quantity of interest, Δ, which is not model-specific, given the data D, is obtained through mixing the model-specific posterior distribution from each individual model:

$$p(\Delta|D) = \sum_{k=1}^{2^q} p(\Delta|M_k, D) p(M_k|D) \tag{6}$$

where p (Δ| M_k, D) is the posterior distribution of Δ given a particular model M_k.

Taking into consideration the above framework, the Bayesian model average estimator for the slope parameters is determined by the posterior mean, defined by:

$$E\left[\beta_j^{BMA}\right] = \sum_{\kappa=1}^{2^q} p(\Delta|M_\kappa, \Delta) p(M_\kappa|D) \tag{7}$$

where $\beta_j^{(k)}$ is the posterior mean under model M_k and equals zero if X_j is not included in the model M_k[6]. That is, the posterior mean is the weighted-average of the model-specific posterior means, where the posterior model probabilities are employed as weights.

The posterior variance for the Bayesian model average estimator is expressed as:

$$Var\left[\beta_j^{BMA}\right] = \sum_{k=1}^{2^q} \left(Var\left[\beta_j^{BMA}|D, M_\kappa\ \right]\ +\ \beta_j^{(k)} \right) p(\ M_\kappa|D) -\ E\left[\beta_j^{BMA}|D\right]^2 \tag{8}$$

[5] When the prior probability distribution of parameters p (θ_k | Mk) is discrete, then summation replaces integration.

[6] According to Raftery (1993) and Drapper (1995), $\beta_j^{(k)}\ =\ E(\ \beta_j^{(k)}|\ D, M_k)$

where the first term captures the average of the posterior variances within models and the second one captures the variance of the posterior means across models (defined as the weighted average of the squared deviations of the model specific from the model averaged estimates). This implies, that even if accurate estimates are calculated in all candidate models, there might be substantial uncertainty about the slope parameters, if those estimates differ across alternative specifications.

4.2 About model and parameter priors

A pure Bayesian approach addresses model uncertainty, but its implementation rests firmly on solving the common challenge of specifying the priors over models in the model space, p(Mk), and the prior distribution for the parameters of each model p (θ_k | M_k). Elicitation of prior parameters can be extremely critical for the outcome and any differences in the results can be attributed to the use of alternative prior assumptions. It is acknowledged (Fernandez, Ley and Steel (2001b), Kass and Raftery (1995), George (1999a) that posterior model probabilities, which are employed as weights for averaging estimates across models, are sensitive to the specification of priors over model-specific parameters. More detailed discussion about model and parameter priors can be found in Appendix.

5 Gini Regression coefficient

Although Least Squares methodology ranks as one of the most popular practices for estimating the relationship between a set of regressors on the conditional expected value of the dependent variable, it relies on certain assumptions, whose violations might result in non-robust estimates. The Gini regression, introduced by Olkin and Yitzhaki (1992), is proposed as an alternative. Its utilization is justified whenever the investigator wants to relax the traditional assumptions, such as the convenient world of normality and the linearity of the model. The Gini methodology is a rank-based methodology that takes into consideration both the variate values and their ranks and it is based on the Gini Mean Difference (GMD) as a measure of dispersion[7].

Between the at least 14 distinct presentations that exist for GMD, the focus has been on the formula that relies on covariances (Lerman and Yitzhaki (1984))[8]. That is, if F(X) is the cumulative distribution function which is uniformly distributed on [0,1], the GMD is expressed as:

$$G = 4E\{X(F(X) - E[F(X)])\} = 4\,cov[X, F(X)] \tag{9}$$

which is four times the covariance between a random variable X and its cumulative distribution function F(X). Equivalently, the above can be rewritten as:

[7] The GMD as a measure of spread/variability was first initiated by Corrado Gini (1912).

[8] For a complete overview of the Gini methodology, the reader is referred to Yitzhaki and Schechtman (2013).

$$G = \frac{1}{3}\frac{cov[X,F(X)]}{cov[F(X),F(X)]} \tag{10}$$

which equals the one third of the slope of the OLS regression curve of the dependent variable, as a function of the observation's positions in the array, F(X), having arrayed the observations in ascending order. Alternatively, it is the weighted average of the slopes defined between two adjacent observations.

There are two types of Gini regressions related to GMD. The first one, known as the R-regression (Hettmansperger (1984)), is based on the minimization of the GMD of the residuals. The second one, known as the semi-parametric approach (Olkin and Yitzhaki (1992)), imitates the OLS methodology by replacing the variance-based expressions by the equivalent GMD terms[9]. The combination of these two allows one to identify and test whether the implicit assumptions about the underlying distributions are supported by the data or not. Apart from that, the superiority of the Gini-based analyses also lies on the fact that Gini estimators are robust to the existence of extreme values or measurement errors and to the asymmetry of the distribution. Under that case, heavy tailed distributions can be used (Serfling (2010)). In addition, only the first moment conditions are needed for the Gini methodology to be implemented (Stuart and Ord (1987, p.58)). Focusing on the second approach, and assuming a simple regression, the population semi-parametric Gini regression coefficient is based on the covariance presentation of the GMD and is obtained by replacing the covariance expressions in the OLS regression coefficient by the corresponding Gini covariances:

$$\beta^N = \frac{cov[Y,F(X)]}{cov[X,F(X)]} \tag{11}$$

where F(X)=R(X) represents the regressor's rank[10].
For the case of multiple Gini regression coefficients, a set of linear equations composed of simple Gini regression coefficients must be solved. Starting from the multiple regression model, expressed in population parameters:

$$Y = a + \beta_1 X_1 + .. + \beta_K X_K + \varepsilon \tag{12}$$

and defining K random variables $R(X_1)$, $R(X_2)$,..,$R(X_k)$, the following identities hold:

$$cov(Y,R(X_1)) = \beta_1 cov(X_1,R(X_1)) + .. + \beta_K cov(X_K,R(X_1)) + cov(\varepsilon,R(X_1)) \tag{13}$$

[9] The Gini estimator taken by the second approach cannot be characterized as "the best" because it is not derived by solving a minimization problem. In contrast, the one derived by the first approach is optimal but it does not have an explicit presentation and it is only expressed numerically.

[10] Empirically the regressor's rank R_x is computed by the formula $R(X) = \frac{\sum_{i=1}^{N} R(X_i)}{N}$.

$$cov(Y, R(X_2)) = \beta_1 cov(X_1, R(X_2)) + .. + \beta_K cov(X_K, R(X_2)) + cov(\varepsilon, R(X_2))$$

$$\vdots$$

$$cov(Y, R(X_K)) = \beta_1 cov(X_1, R(X_K)) + .. + \beta_K cov(X_K, R(X_K)) + cov(\varepsilon, R(X_K))$$

Setting

$$\beta_{\varepsilon j} = \frac{cov(\varepsilon, R(X_j))}{cov(X_j, R(X_j))}, \quad \beta_{kj} = \frac{cov(X_k, R(X_j))}{cov(X_j, R(X_j))}, \quad \beta_{0j} = \frac{cov(Y, R(X_j))}{cov(X_j, R(X_j))} \tag{14}$$

with k,j=1,2,…,K and dividing the three last equations by, respectively, cov(X_1,R(X_1)), cov(X_2,R(X_2)), and cov(X_k,R(X_k)), under the assumption that cov(X_k,R(X_k)) $\neq$0, (k= 1,2,..,K), yields:

$$\beta_{01} = \beta_1 1 + .. + \beta_K \beta_{K1} + \beta_{\varepsilon 1} \tag{15}$$

$$\beta_{02} = \beta_1 \beta_{12} + .. + \beta_K \beta_{K2} + \beta_{\varepsilon 2}$$

$$\vdots$$

$$\beta_{0K} = \beta_1 \beta_{1K} + .. + \beta_K \beta_{KK} + \beta_{\varepsilon K}$$

where the index 0 illustrates the dependent variable, $\beta_{\varepsilon j}$ and $\beta_{kj}$are the regression coefficients in the simple regressions of X_k on R(X_k) and β_{0j} are the semi-parametric Gini regression coefficients as given in presentation (11).

Rearranging, defining the following column vectors β_0= (β_{01} , β_{02},…, β_{0K}), β_ε= ($\beta_{\varepsilon 1}$, $\beta_{\varepsilon 2}$,.., $\beta_{\varepsilon k}$) and provided that the rank of the matrix that consists of the β_k's coefficients is K, it comes:

$$\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} 1 & \beta_{21} & \cdots & \beta_{K1} \\ \vdots & \vdots & & \vdots \\ \beta_{K1} & \cdots & & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{01} & - & \beta_{\varepsilon 1} \\ & \vdots & \\ \beta_{0K} & - & \beta_{\varepsilon K} \end{bmatrix} = A^{-1} [\beta_0 - \beta_\varepsilon] \tag{16}$$

where A^{-1} is a K x K matrix while the β's are K x 1 vectors.

Imposing the set of restrictions, known as "orthogonality conditions", described by:

$$\beta_{\varepsilon k} = 0, \; for\; k = 1,2,..,K \tag{17}$$

the identities of (16) turn into:

$$\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} 1 & \beta_{21} & \cdots & \beta_{K1} \\ \vdots & \vdots & & \vdots \\ \beta_{K1} & \cdots & & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{0K} \end{bmatrix} = A^{-1}\beta_0 \qquad (18)$$

or equivalently

$$\beta^{GINI} = \begin{bmatrix} 1 & \beta_{21} & \cdots & \beta_{K1} \\ \vdots & \vdots & & \vdots \\ \beta_{K1} & \cdots & & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{01} \\ \vdots \\ \beta_{0K} \end{bmatrix} = A^{-1}\beta_0$$

The previous expression shows the Gini estimator β^{GINI} is a function of slope coefficients of semi-parametric simple Gini regressions β_0. Consequently, it is a semi-parametric Gini estimator.
Since, most of the concepts and parameters in the Gini framework are parallel in structure to the OLS framework, it is natural to view presentation (11) as an OLS instrumental variable (IV), where F(X)=R(X) serves as an instrument[11]. According to Olkin and Yitzhaki (1992), when the model is given by:

$$Y = X\, B^{GINI} + \varepsilon \qquad (19)$$

where Y is the dependent variable (N x 1 vector), X $\equiv [x_{ik}]$ is the matrix of regressors (N x K with the first column of ones), B^{GINI} is the vector of parameters to be estimated (K x 1) and ε is the vector of errors (N x1), the semi-parametric Gini regression yields an estimator of β^{GINI},

$$\beta^{GINI} = (R_x'\, X)^{-1}\, R_x'\, \mathrm{Y} \qquad (20)$$

where R_x = R(X).
The Gini regression methodology has been used within different contexts. For instance, Serfling (2010) applies the Gini estimator into the framework of AR (1) models and compares the results to the OLS ones. Similarly, Shelef and Schechtman (2011) use the semi-parametric Gini regression in AR processes, exploiting the fact that two Gini- autocorrelation functions, and consequently two Gini autoregressive coefficients, for each pair of variables exist. Carcea and Serfling (2015) estimate the parameter of ARMA processes under the presence of heavy tailed distributions adopting Gini autocovariance functions.

More recently, Mussard and Ndiaye (2018) consider the semi-parametric Gini methodology into the context of VAR models. Three unbiased semi-parametric Gini estimators are proposed:

[11] Although the Gini regression framework can imitate the OLS concepts, they differ in interpretations and properties. Under the Gini IV analysis, the concept of IV is applied twice. As a first step, the sample's empirical cumulative distribution function replaces the variable, without questioning the validity of the rank to serve as an IV. As a next step, another variable that satisfies all the conditions from an IV perspective is required (Yitzhaki and Schechtman (2004)).

the standard, the Generalized and one derived after relaxing the linearity assumption. Under the presence of outlying observations, all these estimators are proved to be robust compared to the existing estimators (GLS, GMM, ML). Charpentier, Ka, Mussard and Ndiaye (2019) suggest a new Aitken-Gini estimator which, under the presence of heteroskedasticity, is superior compared with the Generalized Least Squared estimator. Further, a Feasible Generalized Gini regression (FGGR) and a Gini-White test are presented where the latter performs better than the usual White-test under the presence of outliers in the data.

6 Data

6.1 Analyzing PISA DATA

PISA (or alternatively, Programme of International Student Assessment) is created by the Organization for Economic Co-operation and Development (OECD) to test the 15-year-old students' skills and knowledge in reading, mathematics and science. Through this procedure vital information is collected regarding students' ability to compete globally, to collaborate for problem solving, to think critically and creatively. What students know and can do, where all students can succeed and what school life means for students' lives are the three questions that constitute the core of this triennial international education survey. Seventy-nine countries and economies and a sample of 600,000 students among 32 million in total, participate in a two-hour test carried out every three years[12]. Only students between the age of 15 years and 3 months and 16 years and 2 months can engage in this survey, while the choice of schools and students is as broad as possible so that a wider range of different educational backgrounds and abilities to be covered.

The PISA test's content is based on the curricula found across the world, without promoting or imposing anyone of those and neither there is a need to find similar characteristics between them. The goal is to assess countries' performance by providing scores for each subject area while the mean score in each subject can be used to rank the participating economies. Since the test is not designed to evaluate individual students, a considerable number of alternative test forms, covering all aspects of test material, exists and gives the opportunity to obtain a much greater coverage of the content[13]. The results, in each test subject, are scaled to follow normal distributions, with mean around 500 score points and standard deviations around 100 score points for OECD countries. Due to the fact that only a sample of students from each country is tested, the estimates are accompanied with statistical uncertainty, meaning that it is impossible to assign an exact rank to each

[12] PISA programme was first launched in 2000. Every three years the main subject of assessment in PISA changes, moving from reading, to mathematics and finally to science and starting all over again, while for the other two subjects PISA provides a summary assessment.

[13] PISA has a two-stage stratified sampling design. Within each participating country a random sample of schools is selected, and then a random sample of students is selected within each school (OECD, 2009). More detailed information on the recruitment, sampling design, procedures and assessment methods are available in a series of technical reports at https://www.oecd.org/pisa/data/2015-technical-report/

participating country, based on its mean score, but rather to place it within a range of positions (that is, between an upper and a lower rank).

To allow comparisons among countries and to conclude whether performance for each country is improved or not, PISA scores are reported at the same scale over time. Under these circumstances, both year-to-year comparisons are feasible and average trends over three or more assessment years can be calculated. Having this information, PISA identifies effective policies and practices that are implemented in economies that record high performance, or in economies that show significant improvement over time. It further, reinforces the participating members that are willing to engage in similar programs, acknowledging that there is no one education model that fits all members, since different economies share different characteristics.

PISA is established and enforced under the authority of education ministries through the PISA Governing Board, the PISA' s decision-making body. All member countries plus partner countries with Associate status elect representatives to the Board, who are a mix of government officials and staff of research and academic institutions. The Board regulates the PISA's policy priorities and standards for data development, analysis and reporting, and supervises their implementation. PISA's financial support is exclusively derived from direct contributions from the participating economies' education ministries[14].

The PISA database has been widely applied in many different studies since it provides the huge advantage to allow for cross-national comparisons of student performances. For instance, Jerrim, Oliver and Sims (2019) use the 2015 wave PISA data for England to investigate the relationship between students' achievement and the inquiry-based science teaching methodology. Zheng, Tucker-Drob and Briley (2018) examine how the family's and school's economic resources as well as the resources found at national level affect the association between science interest and science knowledge, using the 2015 PISA data. Applying the same dataset, S. Cumberworth and E. Cumberworth (2018) reveal if school socioeconomic composition is more strongly associated with standardized test scores among Lower-Socioeconomic Status students than it is among Higher-Socioeconomic Status students. Based on PISA database 2015, Tang (2018) discloses that immigrant adolescents face lower life satisfaction compared to their non-immigrant counterparts, however this gap declines when specific school and family factors are taken into consideration.

Giambona and Porcu (2018) explain how Italian students' achievement is affected by school size, using PISA 2012 data. Yang and Ham (2017), adopting the same data, demonstrate how a well-established anti-discrimination legislation weakens the relationship between first- and second-generation immigrant students and school truancy. Sholderer (2017) utilize PISA 2015 database to explore the impact of social capital on the association between school autonomy and its performance. To inspect how intelligence (IQ) relates to happiness inequality and crime rates, Nikolaev and Salahodjaev (2016) and Burhan et. al. (2014), respectively, use two different

[14] More information about PISA is available at http://www.oecd.org/pisa/

databases whose common denominator is that the construction of IQ variable is based on distinct sources including PISA database. E. Erdogdu and F.Edrogdu (2015) show how home and school environment, access to ICT and student background influence students' academic achievements.

The present study is based on PISA 2015 data, when science was the subject of focus, and uses country level analysis. In this most recent cycle, the 2015 wave of PISA test, 72 participating countries were included for regional comparisons, but some countries are dropped in later analysis due to lacking data. Table 2 in Appendix lists all countries. The variable of interest is proxied by the Students' performance in science. It is expressed in terms of mean score. Indicatively, Students' performance in science ranges from 331,639 to 555,575 score points. The available number of variables in this dataset that can be used as regressors are 167 in total and are classified into 14 categories. However, due to the limited number of observations, Principal Component Analysis is applied to reduce the dimensionality of the dataset, while retaining as much as possible of the variation present in it. The result is a new transformed set of variables, 43 in total, named as principal components, which are uncorrelated, and which are ordered so the first few retain most of the variation present in all the original variables (Jolliffe, 2002). Descriptive statistics for all the OECD variables used in this study are presented in Table 3 in the Appendix.

6.2 Principal Component Analysis (PCA) - Technique

The derivation of Principal Components (PCs) is based on the eigenvectors and eigenvalues of the covariance and/or correlation matrix. An important drawback of PCA based on covariance matrix is that PCs are sensitive to the units of measurement used for the regressors. Thus, before applying the dimensional reduction procedure for multivariate data to identify a small subset of the original variables in the PISA 2015 dataset, all the regressors are standardized because they come with different units. Under that case, the eigenvalues and eigenvectors of a correlation matrix are identical with those of the corresponding covariance matrix. For the computation of the PCs, STATA 14 and EViews 9 programs are used. To decide how many Principal Components should be retained in order to account for most of the variation in the original variables, four rules are used: Cumulative percentage of total variation, Size of variances of Principal Components (Kaiser's rule), Eigenvalue Difference graph and Eigenvalue Cumulative Proportion graph. A detailed description of these rules is supplied in Table 4. in the Appendix. Table 5 in Appendix provides details regarding the new transformed set of variables, namely the principal components. Finally, to regress the dependent variables on this new set of regressors, all the dependent variables have been centered.

7 Empirical Results

Gini-BMA approach is employed to investigate how pre-primary education and other determinants affect Students' performance, by calculating the weighted average of model specific estimates using Gini estimates[15]. The weights attached to each estimate is identical to posterior model probabilities. As a baseline estimation, a universe of all potential models using 43 covariates is taken into consideration. The results are compared to the OLS ones, taken by default when applying BMA methodology. Being totally agnostic about whether any of these regressors is included in the true model, a prior probability of 0.5 is attached to each one, implying a uniform model prior. Regarding the parameter space, the unit information prior (UIP) is adopted, where the integrated likelihood is proxied by the Schwarz Information Criterion (SIC).

Table 6 in Appendix displays the findings for the BMA analysis for the Students' performance in science. The first column shows the posterior inclusion probability that each of the covariates is included in the truth model, while the second and the third columns present the BMA unconditional posterior mean (PSE) and posterior standard deviation (PSD) for each regressor. The remaining three columns provide, respectively, the same amount of information for the OLS case and for the other dependent variables. A covariate is identified as a "robust" determinant, if the posterior inclusion probability exceeds 50%[16].

Referring to the first panel, of the 43 potential/promising candidate regressors only 13 affect the students' performance in science under the Gini analysis. The results grounded in Gini coefficients, suggest that attendance in pre-primary education (i.e. PC11) is a robust determinant of students' performance in science, with a posterior inclusion probability 81.6%, but with an adverse effect on students' performance. However, this apparent negative result conceals the significant variation in years of pre-primary education. That is strongly reinforced by the finding that the percentage of students who had not attended pre-primary education (i.e. EducOut2) enters with high posterior inclusion probability, 73%, and its marginal effect is substantial as well: on average, an one percent increase in no-attendance in pre-primary education reduces the performance by 10.45 score points. In other words, the apparent negative result of PC11 found earlier reflects the significant trade –offs that may exist between entering pre-primary education at a very early age and missing out on parental care at these very early ages, while it is clear that some pre-primary education is crucial and important, but not at all costs. There is a broad opinion in the sense that early early childhood education interventions provide a cognitively simulating environment that enhance school readiness, academic performance, social integration and long-

[15] An alternative extension for the calculation of the estimates can be found in Kourtellos, Stengos and Tan (2013), Durlauf, Kourtellos and Tan (2012) and Eicher, Lenkoski and Raftery (2009), who incorporate the 2SLS estimator into the BMA methodology.

[16] In the paper " Trade Creation and Diversion Revisited: Accounting for model uncertainty and natural trading partner effects", Eicher, Henn and Papageorgiou (2012), following Kass and Raftery (1995), classified the strength of evidence of a regressors' effect into the following categories, sorted by the PIP: if PIP<50%, there is lack of evidence for the effect, if 50%<PIP<75% there is weak evidence for the effect, if 75%<PIP<95% there is positive evidence for the effect, 95%<PIP<99% there is strong evidence for the effect, if 99%<PIP<100% there is decisive evidence for the effect. These cut-offs form an approximation and are not based strictly in statistical theory.

term skill development (Myers (1995), Entwisle and Alexander (1993), Waldfogel (2002), Brooks-Gunn (2003), Carniero and Heckman (2003)). However, along with the early childhood interventions, many studies have found that home conditions are another crucial determinant of child's educational achievement (Bjorklund and Salvanes,2011). Both Velez et.al (1993) and Wößmann (2005) agree that, apart from preschool attendance, parental involvement and family features are key components in students' performance. A child's development begins within the family and depends on the parents' educational and cultural levels (Wößmann, 2005). Waldfogel and Washbrook (2011, and press b) support that parents that are educated and receive high income, spend more time to prepare their children's reading skills. In contrast, parents with lower income and less education have more possibilities to engage in harsh and incompatible parenting teaching behaviours that may negatively affect child's progress. Becker (1981, 1985) and Becker and Tomes (1986) embrace the theory of family to provide a reasonable justification for the failure of preschool education. Furthermore, the percentage of students who had repeated a grade in primary, lower secondary or upper secondary school (i.e. EducOut1), enters with a posterior inclusion probability, 53.8%, but plays a positive role on the students' performance: increasing EducOut1 by 1% raises the probability that the performance will increase by 4.39 score points. The optimistic results about the effectiveness of grade retention can be supported by other studies (Alexander et. al. (1994), Karweit (1999), Lorence, Dworkin, Toenjes and Hill (2002), Greene and Winters (2004), Jacobs and Lefgren (2004), Eide and Goldhaber (2005), Lorence and Dworkin (2006)).

Additional robust determinants are also identified for Students' performance in science. Difference in science performance between immigrant and non-immigrant students (i.e. PC110b) and the mean ratio between students and classroom characteristics (i.e. PC25a) affect performance distinctly. Both appear with a posterior inclusion probability above 90%, (94% and 93% respectively) but with opposite impacts. While an increase in the PC110b coefficient adversely influences Students' performance, an increase in the mean ratio is highly beneficial. Expectations to work in science-related professional and technical occupations at age 30 (i.e. PC23a) exhibit important and negative effects on Students' performance, with a posterior inclusion probability and a posterior mean equal to 78.8% and -11.63 respectively. In contrast, the relative risk of boys expecting to work in science-related professional and technical occupations at age 30 (i.e. PC33b) enters with an inclusion probability 77% and with a considerable positive coefficient equal to 18.04.

The number of students who are evaluated and assessed in alternative ways (i.e. PC44), the number of students with or without an immigrant background (i.e. PC110a), the number of science teachers who are qualified to teach science (i.e. PC111a) and the average time spent per week learning in regular science and no-science lessons (i.e. PC25b), are four variables that are positively effective in performance, appearing with an posterior inclusion probability that ranges between 61% and 68% and posterior means that take values from 4.66 and 7.40. Although PC25b enters positively, the average time spent after school in studying science and no-science lessons (i.e.

PC17) has a negative impact and a posterior mean equal to 5.87. Surprisingly, the third Principal Component that refers to the category of the number of students who are evaluated and assessed in alternative ways (i.e. PC34) plays a negative role: it enters with an inclusion probability of 56%, and a posterior mean equal to -4.73.

A comparison between the Gini-MA results and the OLS-MA ones, presented in the second panel, suggests that the determinants that are important under Gini analysis are not necessarily similar to the ones that are important under OLS analysis. Of the 43 potential/promising candidate regressors only 10 affect the students' performance in science under OLS. The number of students who refer skipping/arriving late at classes two weeks prior to the test (i.e. PC26c), and the difference in science performance (i.e. PC39c) appear, under this case, to affect performance. Entering with posterior inclusion probabilities 78% and 71%, respectively, with PC26c appear to have a negative effect, while PC39c appears to have a positive one. The percentage of students who had not attend pre-primary education (i.e. EducOut2) continues having a negative impact, although smaller (8.75 score points), with an inclusion probability of 84%. This is similar to the findings of the Gini-MA analysis, yet an important difference is that the effect of pre-primary education is not important with an inclusion probability less than 50% (0.31%).The case of attendance in pre-primary education (i.e. PC11) with the Gini-MA analysis was found to be strongly robust with a negative effect on students' performance and as was argued that apparent negative result may conceal the significant variation in years of pre-primary education and the potential benefits and costs between too many or too few years of pre-primary education. However, this is not the case for the percentage of students who had repeated a grade in primary, lower secondary or upper secondary school (i.e. EducOut1), which now appears with a negative posterior mean equal to -10.257 and much higher inclusion probability,75%. This is the most important difference so far since it implies that repeating a grade does not have a beneficial effect on performance.

The number of students with or without an immigrant background (i.e. PC110a) and the mean ratio between students and classroom characteristics (i.e. PC15a) enjoy strong posterior support for being important explanations for students' performance in science. Both receive an inclusion probability above 90%, and their posterior means are of the same magnitude but of opposite signs: the PC15a coefficient is positive and equal to 16.629, while the PC110a coefficient is negative and equal to 16.699. It certainly appears to be true that expectations to work in science-related professional and technical occupations at age 30 (i.e. PC13a), the relative risk of boys expecting to work in science-related professional and technical occupations at age 30 (i.e. PC23b) and average time spent per week learning in regular science and no-science lessons (i.e. PC15b) contribute to science performance. The results found using Gini-MA approach are confirmed, since, now, these variables enter with higher inclusion probabilities, 89%, 86% and 75% and with posterior means equal to -14.292, 22.123 and 6.324 respectively. In contrast, the number of students who are evaluated and assessed in alternative ways (i.e. PC34) seems to positively affect performance and enters with an inclusion probability equal to 57% and a posterior mean equal to 3.87.

Table 7 summarizes all the robust determinants under the Gini and OLS analysis, respectively. The first four regressors (i.e. the number of students with or without an immigrant background (i.e. PC110a), the number of students who are evaluated and assessed in alternative ways (i.e. PC34), the percentage of students who had repeated a grade in primary, lower secondary or upper secondary school (i.e. EducOut1), percentage of students who had not attended pre-primary education (i.e. EducOut2)) appear to be significant under both analyses. Only the last variable retains the negative coefficient, while for the rest, the sign changes between the two cases. Regarding the next four variables (i.e. mean ratio between students and classroom characteristics (i.e. PC25a, PC15a), the average time spent per week learning in regular science and no-science lessons (i.e. PC25b, PC15b), expectations to work in science-related professional and technical occupations at age 30 (i.e. PC23a, PC13a), relative risk of boys expecting to work in science-related professional and technical occupations at age 30 (i.e. PC33b, PC23b)), all appear to retain their sign but each variable refers to a different principal component, when comparing between the two analyses. The rest five variables that appear in the second column, are robust determinants only under the Gini analysis (i.e. difference in science performance between immigrant and non-immigrant students (i.e. PC110b), attendance in pre-primary education (i.e. PC11), average time spent after school in studying science and no-science lessons (i.e. PC17), the number of students who are evaluated and assessed in alternative ways (i.e. PC44), the number of science teachers who are qualified to teach science (i.e. PC111a)), while the rest two variables that appear in the fourth column are robust determinants only under the OLS analysis (i.e. difference in science performance (i.e. PC39c), the number of students who refer skipping/arriving late at classes two weeks prior to the test (i.e. PC26c)).

Image plots in figures 1a. and 1b. in Appendix demonstrate the sign and the importance of the regressors in the universe of models. These graphical representations highlight how the estimated coefficients fluctuate for the top 100 models shown in the horizontal axis, scaled by their posterior model probability. Blue and red colors indicate the inclusion of the regressor with a positive and negative posterior mean respectively. White color indicates non-inclusion (or a zero coefficient). Robust covariates retain the same sign pretty much throughout the model space.

The top panels of figures 2a. and 2b. in Appendix provide the distribution of model sizes for the baseline exercise for Students' performance in science. With 2^K possible variable combinations, a uniform model prior implies a common prior model probability of 21.5. However, the graphs show a posterior model size distribution equal to 19.3983 and 19.0188, for the Gini and OLS case respectively. The low panels of figures 2a. and 2b. demonstrate the best 100 models encountered, ordered by their analytical posterior inclusion probability (red line) and plot their MCMC iteration counts (blue line)[17]. At 0.9308 and 0.83, these correlations are far from perfect but the differences from an exact likelihood approach are practically indiscernible and already indicate a satisfactory rate of convergence.

8 Conclusions

This paper identifies the robust determinants of students' performance in Science. To ensure a comprehensive search, the analysis accounts for a rich set of possible regressors by employing Gini-BMA methodology. This approach constructs estimates that do not depend on a particular model specification, but rather they are conditional on the model space. A weighted average of Gini and OLS coefficients are calculated, respectively, where the weights are given by the posterior model probabilities.

Once model uncertainty is accounted for, the results can be summarized as follows. First, of the 43-promising candidate regressors only 13 affect the students' performance in science under the Gini analysis, while only 10 under OLS. Among the factors that are robust under both cases are: the percentage of students who had not attend pre-primary education (EducOut2), the percentage of students who had repeated a grade in primary, lower secondary or upper secondary school, the mean ratio between students and classroom characteristics, the expectations to work in science-related professional and technical occupations at age 30, the relative risk of boys expecting to work in science-related professional and technical occupations at age 30, the average time spent after school in studying science and no-science lessons and the number of students who are evaluated and assessed in alternative ways. However, in most cases, differences are observed in the signs of the coefficients.

Second, only under the OLS case, the number of students who refer skipping/arriving late at classes two weeks prior to the test, the difference in science performance and the number of students with or without an immigrant background are considered important for the performance in science. In contrast, only under the Gini case, the number of students who are evaluated and assessed in alternative ways, the number of students with or without an immigrant background, the number of science teachers who are qualified to teach science, the difference in science performance between immigrant and non-immigrant students, the average time spent per week learning in regular science and no-science lessons, and the attendance in pre-primary education affect performance.

9 Appendix

9.1 Appendix A-Model Priors

[17] The model space is constructed by using birth-death MCMC sampler based on 10^6 burn-ins and $5x10^6$ draws.

Beginning with considerations for choosing model priors, $p(M_1),\ldots,p(M_k)$, the most common approach is an non-informative prior which favors all candidate models equally. For a model with p independently included regressors and size Ξ, the model size follows a Binomial distribution with probability of success ξ:

$$\Xi \sim \text{Bin}\,(\text{p}, \xi) \tag{A.1}$$

where p is the total number of candidate regressors and ξ is the prior inclusion probability for each variable. Based on the above, the prior of a model M_k with p_k regressors is described by:

$$p\,(M_k) \;=\; \xi^{q_k}(1\;-\;\xi)^{q-\,q_k} \tag{A.2}$$

Raftery (1988), Raftery et al. (1997), Fernandez et al. (2001 a,b) and George and McCulloch (1993), fix ξ to equal 0.5 so that every regressor has the same a priory probability[18] . This leads to the uniform model prior, which can be considered as a benchmark, that assigns equal prior probability to all models, implying that $p(M_k)= 2^{-p}$ for each p and that expected model size is p/2. Mitchell and Beauchamp (1988) introduced the more general model prior structure, when prior information about the relevance of a variable is applicable, namely:

$$p\,(M_k) \;=\; \prod_{j=1}^{q} \pi_j^{\delta_{k_j}}(1\;-\;\pi_j)^{1\,-\,\delta_{k_j}} \tag{A.3}$$

where π_j ε [0 ,1] is the prior probability that variable X_j is included in the model and δ_{kj} =1 if X_j is included in M_k and 0 otherwise. Usually, it is assumed that $\pi_j = \pi$ for j= 1,….,p. For π=0.5, (A.3) corresponds to a uniform prior across model space, while π<0.5 imposes a penalty for large models. Assigning π_j =1, guarantees the inclusion of variable j in all candidate models[19]. An extension on this approach was proposed by Brown et al (1998;2002) and Ley and Steel (2009), who assign a hyperprior on the probability of inclusion, π, converting it into a random variable drawn from a Beta distribution[20].

When little information is available about the relative validity of the candidate models, assuming independent inclusion of regressors a priori seems a "neutral" choice[21] . However, it might be deceptive/contentious/quarrelsome in some circumstances (Chipman et al. (2001)). The

[18] Sala-i-Martin et al. (2004) perform sensitivity analysis to examine how different values for ξ affect their results. They pre-specify a prior mean model size, kbar, implying that each variable has a prior probability of inclusion equal to kbar/K, with K being the total number of potential regressors. As a special case, when kbar=K/2, equal probability is assigned to each possible model.

[19] George and McCulloch (1993), Volinsky et al. (1997), and Madigan and Raftery (1994) apply this approach in the context of linear regressions, Cox models, and graphical models respectively.

[20] In that case, and according to Bernardo and Smith (1994, p.117), the prior on model size is a Binomial-Beta distribution.

[21] Both the Binomial and the Binomial-Beta priors are based on this assumption.

uniform model prior does not take into consideration interrelations between different variables, replicating a problem comparable to the irrelevance of independent alternatives (IIA) in the discrete-choice literature[22]. When the goal is to evaluate the relative significance of distinct theories and to define non-informative model priors across theories, the uniform prior is inappropriate, since the researcher can change the prior weights across theories simply by introducing “redundant” proxy variables for each theory.

George (1999b) proposed a dilution prior as a solution to the interrelations between variables. If the set of candidate regressors includes variables that represent the same concept, George’s dilution prior increases the prior probabilities of models not containing these correlated predictors. However, this is not always the case, since variables are often measures of different ideas but are still correlated. Under this condition, the straightforward use of this prior penalizes larger models.

To deal with the interdependencies across theories due to the addition of “redundant” variables, Durlauf, Kourtellos and Tan (2007;2012) choose the prior probability that a particular theory - defined as the set of variables that are used as proxies for that theory - is included in the true model to equal 0.5. This assumption captures the non-informativeness (i.e. agnosticism) across theories but also ensures that the probability of inclusion of one theory in a model does not exclude other theories from being relevant. The question now is how to assign prior probabilities across the set of variables within each theory. To answer this, specification uncertainty should be taken into consideration. This problem is related to existence of correlations between potentially unrelated proxy regressors within theories. To handle this, they introduce a modified version of George’s (1999b) dilution prior. Selecting a theory T, with p_T regressors, as a priori proper, they construct a binary vector γ_T for each possible combination of these p_T proxies. The conditional prior probability assigned to each γ_T is given by:

$$\mu^D(\gamma_T) = |R_{\gamma_T}| \prod_{j=1}^{q_T} \pi_j^{\gamma_i}(1 - \pi_j)^{1-\gamma_j} \tag{A.4}$$

where π_j is the prior inclusion probability of each proxy variable in theory T, which is equal to π_j=0.5 for j = 1,….,p_T and $|R_{\gamma_T}|$ is the correlation matrix for the set of variables included in the binary vector γ_T. When regressors are collinear, $|R_{\gamma_T}|$ takes the value of zero, whereas when regressors are orthogonal, it is equal to 1. This structure penalizes models that include irrelevant variables and retains weights on informative models.

A similar approach can be found in Brock and Durlauf (2001) and Brock et al. (2003) who focus on economic theories rather than individual regressors and use hierarchical tree structure to construct model priors. A related idea was expressed by Chipman et al. (2001), who assigns probability to neighborhoods of similar models. More recently, George (2010) develops dilution model priors classified in three distinct approaches: the tessellation defined dilution priors[23], the

[22] See Brock ad Durlauf (2001), and Brock, Durlauf and West (2003) for further analysis.

[23] Moser and Hofmarcher (2014) provide an extended analysis on the implementation of this approach.

collinearity adjusted ones and the model distance based. The key characteristic of these is to assign prior probabilities more uniformly across neighborhoods of models rather across models[24]. Despite it seems a sufficient answer to the dilution property, this prior structure obliges/requires a decision on which proxies are classified under a specific theory and which models belong to the same neighborhood. Such decisions are not within reach in most of the cases.

9.2 Parameter priors/Prior distributions of parameters

In the context of Bayesian framework, to complete the Normal linear regression model described in (2), a prior distribution for the parameters $\theta_k = (\beta_0, \beta^{(k)}, \sigma)$ is needed. This distribution will be given through a density function:

$$p\ (\beta_o, \beta^{(k)}, \sigma \mid M_k) \tag{A.5}$$

which consists a key component in the marginal or integrated likelihood of model M_k:

$$p\ (D \mid M_k) = \int p(D \mid \beta_o, \beta^{(k)}, \sigma, M_k)\ p(\beta_o, \beta^{(k)}, \sigma \mid M_k) d\beta_o d\beta^{(k)} d\sigma \tag{A.6}$$

and affects right away the posterior model probabilities $p\ (M_k \mid D)$. Two challenging questions arise regarding the computation of $p\ (M_k \mid D)$ and the influence of the assumptions made for prior distributions on the latter quantity. Analytical answer to the first one is provided in the next section. Apart from purely computational features, the choice of “rational” prior parameter distributions remains unresolved and depends mainly on the availability of prior information. When information about the parameters is given, informative priors can be constructed (e.g. Jackman and Western (1994)). However, under little or absence of prior information, choosing a distribution/density for (13) is a very complex task. Consequently, many efforts have been made to establish “default priors” or “reference priors” that can be applied in all such cases.

9.2.1 A non-informative prior for the intercept and for σ

Following Fernandez et al. (2001b), “non-informative” improper priors have been adopted for the common intercept and the scale σ, such

$$p(\beta_o) \propto 1 \tag{A.7}$$

$$p(\sigma) \propto \sigma^{-1} \tag{A.8}$$

[24] Another promising approach to dilution prior construction is suggested by Garthwaite and Mubwandarikwa (2010), who construct prior model weights using the correlation matrix between models. This matrix reflects the similarities between models and assigns small weights to those who are highly correlated.

assuming common prior distribution for σ across models and that β_0 is independent of $\beta^{(k)}$ and σ, so that p (β_0, $\beta^{(k)}$, σ) = p(β_o)p($\beta^{(k)}$| σ^2)p(σ^2).

9.2.2 Informative priors for the regression coefficients

In general, direct use of improper noninformative priors for model-specific parameters is not allowable because their arbitrary norming constants remain in the integrated likelihoods and lead to uncertain model probabilities[25](Jeffreys (1961); Berger and Pericchi (2001)). Conventional proper priors for regression coefficients have been relied on the natural-conjugate approach, which assigns a conditional normal prior on the k-th model's parameter ($\beta^{(k)}$ | σ^2) with zero mean and the variance proposed by Zellner (1986), leading to the following prior distribution (Fernandez et al. (2001b)):

$$p(\beta_o, \beta^{(j)}, \sigma \mid M_j) \;\propto\; \sigma^{-1} f_N^{k_j}(\beta^{(j)} \mid 0, \sigma^2 (g Z_j' Z_k)^{-1}) \qquad \text{(A.9)}$$

where f_N^q (w | m, V) denotes the density function of a q-dimensional Normal-distribution of w with mean m and covariance matrix V, and g is a scalar that measures how important are the prior beliefs of the researcher[26]. The above prior distribution allows for exclusion of regressors from some of the models, represented by a prior point mass at zero (known as the "spike-and-slab" approach in Mitchell and Beauchamp (1988)).

Efficiently, the thorny problem of picking a prior distribution for β can be solved only by selecting a single parameter g. Under this "benchmark" prior structure, incorporating subjective prior knowledge into the analysis is not feasible or desirable, resulting in little influence on posterior inference. These "automatic" priors depend only on the number of independent variables and the sample size (Fernandez et al. (2001b)). In the same spirit, Kass and Wasserman (1995 and Raftery (1995) recommend "unit information" priors, which contain the same amount of information as a regular single observation. In contrast, Raftery et al. (1997) display "weakly informative" proper priors, which are data dependent through the response variable[27].

To demonstrate the behavior of several notorious priors that belong to the above-mentioned categories (refer to the priors 1-11 and 14, presented in Table1 in Appendix), Eicher et all. (2007) compare their predictive performance by employing growth and simulated data and conclude that Unit information Prior in combination with uniform model prior surpass all the rest. In contrast,

[25] To overcome this problem, Berger and Pericchi (1996) and O'Hagan (1995) apply intrinsic Bayes factors and fractional Bayes factors, respectively, but their approaches suffer from inconsistencies.

[26] A large value of g implies a small true model, that is, many of the regressors equal zero, while a small g supports the existence of a large model. When g□0, the β estimator is the Least Squares estimator of the full/ "kitchen sink" model (George and Foster (2000)-Calibration and empirical Bayes variable selection).

[27] Tobias and Li (2004) apply this prior choosing the following values for the parameters: $V=2.85^2 I_k$, $\lambda=0.28$, $v=2.58$, $\mu_\beta= 0$.

Ley and Steel (2009) recommend the avoidance of UIP in the context of growth regressions or under the presence of large number of potential regressors, proposing instead the use of prior $g=1/k^2$ (combined with the assumption that the inclusion of each regressor is independent and equal to 0.5).

According to Liang et al. (2008), when fixed g priors are used to construct Bayes factors, the result might suffer from the Bartlett's and the Information paradoxes. To overcome this complication, they investigate fully Bayes approaches and suggest three alternatives: Global and Local empirical Bayes procedures, the multivariate Zellner-Siow Cauchy priors (initially introduced by Zellner and Siow (1980)) and a family of prior-probabilities imposed on g[28]. Because this hyper-g prior family for g lacks model selection consistency, they provide a modification, known as hyper-g/n prior family[29]. In the same strand of literature, Ley and Steel (2012) attach a proper hyperprior to g, which corresponds to a shrinkage factor $\delta = g/(1+g)$ that follows a Beta prior distribution. This "benchmark" Beta prior with c=0.1 and a hyper-g/n prior with $\alpha=3$ (and $\alpha=4$), is compared with existing priors[30] in terms of model selection consistency, avoidance of information paradox and empirical behavior/performance.

Feldkircher and Zeugner (2009) finalize the analysis presented by Liang et al. (2008), by adding the posterior distribution of $\beta|y,X$, its second moments, and the second moment of the shrinkage factor. Regarding the computation of these posterior expressions, they use algebraic transformations and implement accurate statistics to overcome possible errors that occur when Laplace approximations are applied. In a simulation study with noisy data, they show that hyper g-prior spreads the posterior mass more evenly among the candidate models compared to the "Benchmark prior" (FLS (2001b)).

A completely different approach, known as the Bayesian Averaging of Classical Estimates (BACE) methodology, is given by Sala-i-Martin et al. (2004), whose analysis is not based on g-parameter priors, but rather information is extracted from the data and the final estimates are the result of averaging OLS estimates across models. The weights assigned to each model is the logarithm of the likelihood function, agreeing to Schwarz model selection criterion[31].

[28] Strawderman (1971)-Proper Bayes minimax estimators of the multivariate normal mean- and Cui and George (2007)-Empirical Bayes vs Fully Bayes variable selection- study priors that belong to this family of prior probabilities for g.

[29] According to Liang et al. (2008), the three alternative solutions resolve the information paradox and are consistent under prediction. However, only Zellner-Siow priors are consistent for model selection.

[30] These are Zellner and Siow (1980), Maruyama-George (2011), Bottolo-Richardson (2008), Feldkircher and Zeugner (2009), Liang et al. (2008), Carvalho et al. (2010), Forte et al. (2010).

[31] Although Sala-i-Martin et al. support that "BACE limits the effect of prior information", it is important to mention that even if prior assumptions are implicit now, this does not imply that the dependence on the them becomes less important (Ley and Steel (2009)).

9.2 Appendix B-Figures

Figure 1: Model Inclusion Probability on Best 300 Models for the OLS case (right) and Gini case(left)

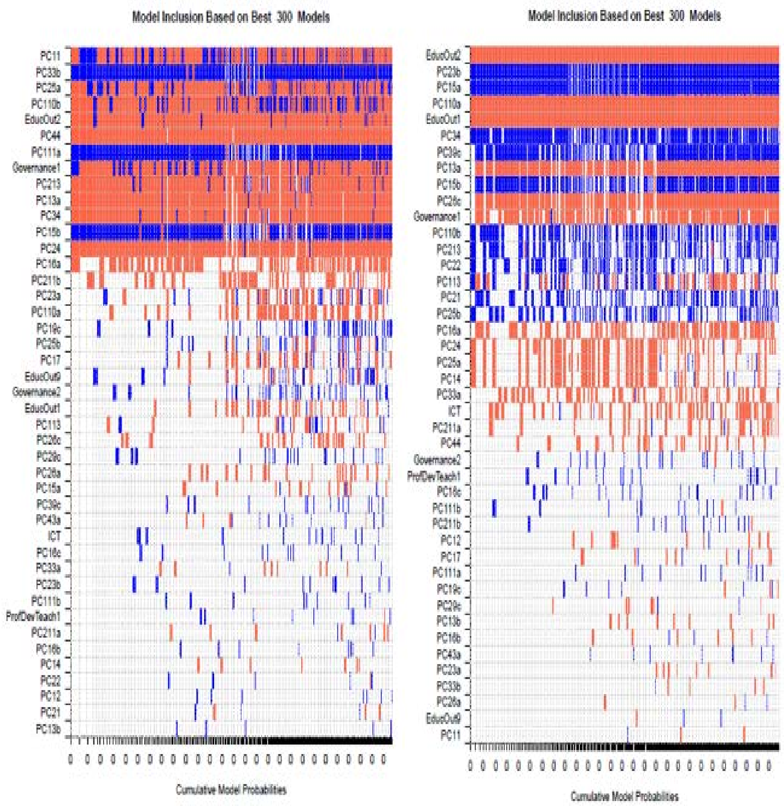
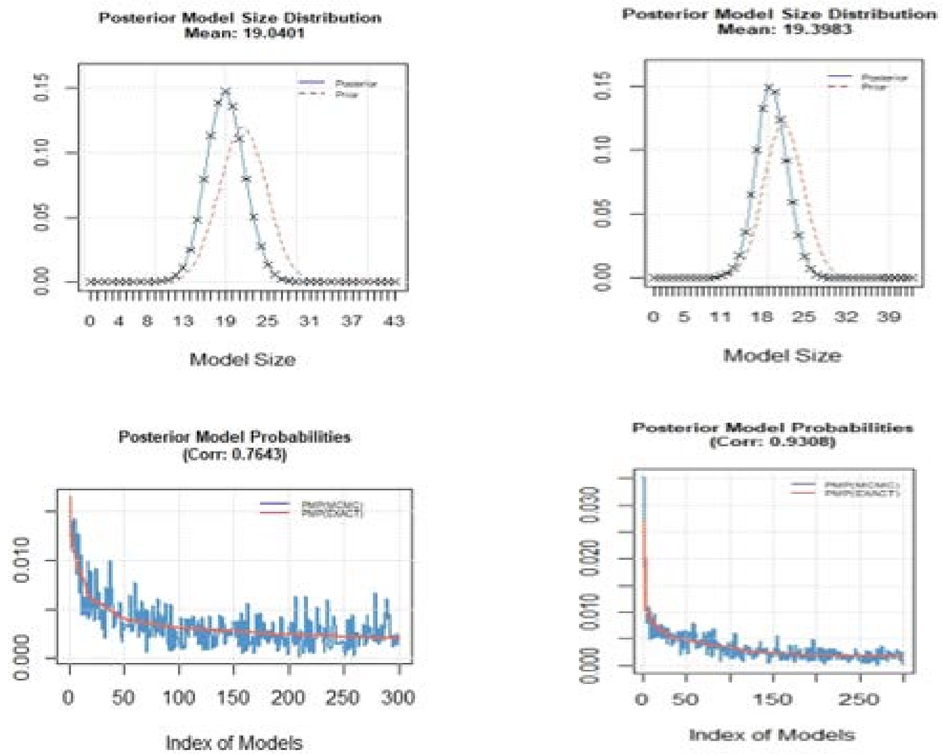


Figure 2: Posterior Model Size Distribution and Posterior Model Probabilities for the OLS case (left) and Gini case(right)



9.3 Appendix C-Table

Table 1: Parameter prior structures

Automatic Priors (Fernandez et al. (2001b) (FLS))		
G-prior		Description
1) $g_{oj} = \frac{1}{n}$	The log Bayes factor obtained under this prior behaves asymptotically like the Schwarz criterion. This prior leads to consistency.	It assigns the same amount of information as a regular single observation. It is like "unit information prior" proposed by Kass and Wasserman (1995), but with zero mean instead of MLE.
2) $g_{oi} = \frac{k_j}{n}$	The log Bayes factor obtained under this prior behaves asymptotically like the Schwarz criterion. This prior leads to consistency.	As more independent variables are added in the model, more information is attached to this prior (i.e. the discrete point mass at zero for β shrinks.)
3) $g_{oi} = \frac{k^{\frac{1}{k_j}}}{n}$	The log Bayes factor obtained under this prior behaves asymptotically like the Schwarz criterion. This prior leads to consistency.	As more independent variables are added in the model, less information is attached to this prior.
4) $g_{oi} = \sqrt{\frac{1}{n}}$	The penalty applied for selecting larger models is smaller compared to the Schwarz (BIC) criterion. This prior leads to consistency.	It is an in-between case of prior 1 and attributes smaller asymptotic penalty for selecting larger models.
5) $g_{oi} = \sqrt{\frac{k_j}{n}}$	The penalty applied for selecting larger models is smaller compared to the Schwarz (BIC) criterion. This prior leads to consistency.	It is an in-between case of prior 2, more information is attached to this prior as the number of regressors increases (i.e. the

		discrete point mass at zero for β shrinks).
6) $g_{oi} = \frac{1}{(Ln(n))^3}$	This prior leads to consistency.	As n becomes large, this prior mimics the Hannan-Quinn (1979) criterion with C_{HQ}=3.
7) $g_{oi} = \frac{Ln(k_j + 1)}{Ln(n)}$	This prior leads to consistency.	As n becomes larger, this prior decreases even slower and has asymptotic convergence to the Hannan-Quinn (1979) criterion with C_{HQ}=1.
8) $g_{oi} = \frac{\delta\gamma^{1/k_j}}{(1-\delta\gamma^{1/k_j})}$	This prior does not lead to consistency in general.	This natural conjugate prior structure is proposed by Laud and Ibrahim (1996) and is subjectively elicited through predictive implications. The suggested values for the parameters are: γ<1, such that the prior increases when more regressors k_j are added in the model, the shrinkage factor δ = $g_{0j}/(1+g_{0j})$ belongs to the interval [0.10,0.15] (i.e. the weight of the " prior prediction error" in the Bayes factors) and the number of regressors k_j ranges between 1 and 15. To cover this interval, FLS choose the values γ=0.65 and δ=0.15

9) $g_{oi} = \frac{1}{k^2}$	This prior does not lead to consistency in general.	It coincides with the Risk Inflation Criterion proposed by Foster and George (1994)
10) $g_{oi} = \frac{1}{max\{n, k^2\}}$	This prior combines the consistent Bayes factors shaped under prior 1 with the remarkable small sample performance of prior 9.	It is a combination of prior 1 and prior 9 and is the most preferred prior for FLS (2001b)
Unit Information prior (Kass and Wasserman (1995); Raftery (1995))		
11) $g = N$ (g-UIP)	This prior is known as g-UIP. The posterior model probabilities (and thus the Bayes factors) are approximated by the Schwarz criterion (BIC). That is, log pr $(D\|M_k) \approx c - 1/2BIC_k$ where $BIC_k = n\log(1-R_k^2) + p_k\log(n)$ c is a constant that does not change across models The g-UIP prior does not resolve the information paradox.	It assigns the same amount of information as a regular single observation.
Foster and George (1994)		
12) $g = q^2$ (g - RIC)	This prior is known as g-RIC. The g-RIC prior does not resolve the information paradox.	It is related to Risk inflation Criterion (RIC)
13) $g = max\{N, q^2\}$ (g-BRIC)	This prior is known as the "Benchmark prior" and it is called the "g-BRIC". Regarding predictive performance, it is the most favorite by FLS (2001b). The g-BRIC prior does not resolve the information paradox.	It is a combination of prior 11 and prior 12.
"Weakly-informative" priors or equivalently "Data-Dependent" priors (Raftery, Madigan and Hoeting (1997))		
14) $B \sim N(\mu, \sigma^2 V)$ where $V = \sigma^2\varphi^2(\frac{1}{nX'X})^{-1}\frac{u\lambda}{\sigma^2} \sim \chi^2$	This prior belongs to standard normal gamma conjugate class of priors. The hyperparameters to be selected are: v, λ, μ (a (p+1) vector) and V (a (p+1)x(p+1) covariance matrix for β referring to model M_k) The marginal likelihood for Y is:	It is defined by four hyperparameters: if the full model has an R^2 less than 0.9 then φ=0.85, v = 2.58, λ= 0.28. In contrast, if the full model has an R^2

	$p\ (Y \mid \mu_i, V_i, x_i, M_i) = \frac{\Gamma\ (\frac{v\ +\ n}{2})(v\lambda)^{v/2}}{\pi^{\frac{n}{2}}\ \Gamma\ (\frac{v}{2}) \lvert I\ + X_i v_i x_i^t \rvert^{\frac{1}{2}}}\ x$ $x\ [\lambda v + (Y - x_i\mu_i)^t x(I\ +\ x_i v_i x_i^t)^{-1}(Y - x_i\mu_i)]^{-(v+n)/2}$ The Bayes factor for model M_o versus model M_1 is: $B_{oi} =\ (\frac{\lvert I\ +\ x_1 v_1 x_1^t \rvert}{\lvert I\ + x_0 v_0 x_0^t \rvert})^{1/2} (\frac{a_0}{a_1})^{-(v+2)/2}$	more than 0.9 then φ=9.2, v = 0.2, λ= 0.1684.
Hyper -g priors (Liang, Paulo, Molina, Clyde and Berger (2008))		
15) $\pi(g)\ =\ \frac{(n/2)^{1/2}}{\Gamma(1/2)}\ g^{-3/2} e^{-n/2g}$	To analyze the properties of priors on g, a quantity, named as shrinkage factor, is used and defined as: $\delta\ =\ \frac{g}{(1\ +\ g\ )}$ following a distribution: $p(\delta)\ =\ \frac{\sqrt{\frac{n}{2}}}{\Gamma(\frac{1}{2})} \delta^{-3/2}(1-\delta)^{-1/2} exp(-\ \frac{n(1\ -\ \delta)}{2\delta})$ The Zellner- Siow priors are a mixture of g-priors with an Inverse Gamma (1/2, n/2) prior on g. Under the Zellner-Siow prior, there are not closed- form solutions for marginal likelihoods. The Zellner-Siow prior resolves the information paradox and it is asymptotic consistent for prediction and model selection. Usually in the literature, the shrinkage factor follows a Beta prior distribution Beta(b,c), leading the prior on g to follow a Gamma-Gamma distribution (Bernardo and Smith (1994), p. 120) or, alternatively, an inverted Beta distribution (Zellner (1971), p.375): $p(g) = \frac{\Gamma(b+\ c)}{\Gamma(b)\Gamma(c)} g^{b-1}(1\ +\ g)^{-(b+c)}$ This prior on g leads to the following prior on the regression coefficients:	These are the multivariate Cauchy priors, introduced by Zellner and Siow (1980), where the prior on g follows an Inverse Gamma (1/2, n/2) distribution.

	$p(\beta_k \| M_k, \sigma) = \frac{\Gamma(b+c)\Gamma(c+\frac{k_j}{2}) \|Z_j' Z_j\|^{1/2}}{\Gamma(b)\Gamma(c)(2\pi)^{k_j/2}\sigma^{k_j}} x\, \Psi\,(c + \frac{k_j}{2}, \frac{k_j}{2} - b + 1; \frac{\beta_j' Z_j' Z_j \beta_j}{2\sigma^2})$	
16) $\pi(g) = \frac{a-2}{2}(1+g)^{-a/2}$ *with g >0, a>2*	To analyze the properties of priors on g, the shrinkage factor δ = g / (1 + g) is mobilized, which follows a Beta distribution Beta(b,c)=Beta(1, (α/2) -1), α>2. For α=4, it becomes a uniform distribution. Any choice between 2<α≤4 might give reasonable results. Liang et al. (2008) choose the values α=3 and α=4. The posterior distribution of g under model M_k has a closed-form solution given by: $p(g\|M_k) = \frac{p_k + a - 2}{2\,{}_2F_1\left((n-1)/2; 1; (p_k+a)/2; R_k^2\right)} x\, (1+g)^{(n-1-p_k-a)/2}[1+(1-R_k^2)g]^{-(n-1)/2}$ The hyper-g prior derived from this family resolves the information paradox, under the constraint that 2< α ≤3 (when sample size is minimal). Also, it is asymptotic consistent for prediction but not for model selection.	This is a family of hyper-g priors proposed by Liang et al. (2008).
17) $\pi(g) = \frac{a-2}{2n}(1+\frac{g}{n})^{-a/2}$	The shrinkage factor δ= g / (n +g) follows a Beta distribution Beta (b,c)= Beta (1 ,(a/2) – 1) This family of hyper- g/n priors resolves the asymptotic inconsistency faced by the hyper- g priors in 16.	This is a family of modified hyper-g/n priors proposed by Liang et al. (2008).
18) $g_\gamma^{EBL} = max\{F_\gamma - 1, 0\}$ $F_\gamma = \frac{R_\gamma^2 / p_\gamma}{(1-R_\gamma^2) / (n-1-p_\gamma)}$	This g-prior is based on the Local Empirical Bayes approach. This g-prior resolves the information paradox, and is asymptotic consistent for prediction. However, it is not consistent for model selection.	A distinct g-prior is evaluated for each model. The estimate for g coincides with the maximum marginal likelihood estimate, under the restriction to be a positive number.
19)	This g-prior is based on the Global Empirical Bayes approach.	A common g-prior is evaluated for all models. The estimate for g

$g^{EBG} = argmax_{g>0} \sum_{\gamma} p(M_\gamma) \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}}$	There is not closed-form solution for this g-prior since the marginal likelihood is not tractable. However, as suggested by George and Foster (2000), numerical optimization can be used as a solution. Liang et al. (2008) recommend an EM algorithm that provides a maximum likelihood estimation for g : $g^{(i+1)} = max\{\sum_{\gamma} p^{(i)}(M_\gamma \| Y) x \frac{R_\gamma^2 / \sum_{\gamma'} p^{(i)}(M_{\gamma'} \| Y) p_{\gamma'})}{(1-(g^{(i)}/(1+g^{(i)}))R_\gamma^2)/(n-1)} - 1, 0\}$ Under this algorithm, Empirical Bayes posterior model probabilities can be attained at convergence. This g-prior resolves the information paradox, and is asymptotic consistent for prediction. However, it is not consistent for model selection.	coincides with the maximum marginal likelihood of the data, averaged over all models.
20) $\pi_{a,b}(c) = M(1+c)^{-(1+a)} exp\{-\frac{b}{1+c}\}$ where $M = b^a(\int_0^b t^{a-1}e^{-t}dt)^{-1}$ and $\pi_a(c) = a(1+c)^{-(1+a)}$ for cϵ $(0, \infty)$	This g-prior is based on Fully Bayes approaches. The variance is known. To implement this approach, hyperpriors for two hyperparameters c and w should be defined. c is a hyperparameter for the average size of the regression coefficients matrix for model γ (β_γ). w is a hyperparameter for the average number of nonzero coefficients in the model γ. The term (1+c) for cϵ (0, ∞) follows an incomplete Gamma distribution (a, b), where a and b are hyperparameters, with default choices at b= 0 and a=1.	It is suggested by Cui and George (2007).
21) $g(x,\lambda) = K_2 \lambda^{\frac{1}{2}p-a} exp\left(-\frac{1}{2}\lambda\|\|\chi\|\|^2\right),$ $0 < \lambda \leq 1, \|\chi\|^2 > 0$ where p≥6 and 0≤a<1 and if p=5 then 1/2≤a<1	These family of priors are used to provide proper Bayes minimax estimators when the mean of a multivariate normal distribution is estimated (under identity covariance matrix and a loss in squared errors). The unconditional density for λ with respect to Lebesgue measure is given by $\lambda^{-a}/(1-a)$ for any a, 0≤a<1	It is introduced by Strawderman (1971).
Hyper -g priors (Ley and Steel (2012))		
22)	The shrinkage factor δ=g/(1+g) follows a Beta prior distribution Beta(b,c), leading	It is suggested by Ley and Steel

$p(g) = \frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)} g^{b-1}(1+g)^{-(b+c)}$	the prior on g to follow a Gamma-Gamma distribution (Bernardo and Smith (1994), p. 120) or, alternatively, an inverted Beta distribution (Zellner (1971), p.375): $$p(g) = \frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)} g^{b-1}(1+g)^{-(b+c)}$$ which has the following properties: $$E[g] = \frac{b}{c-1}, provided\ that\ c > 1$$ and $$Var[g] = \frac{b(b+c-1)}{(c-1)^2(c-2)}, provided\ that\ c > 2$$ This prior on g leads to the following prior on the regression coefficients: $$p(\beta_k \| M_k, \sigma) = \frac{\Gamma(b+c)\Gamma(c+\frac{k_j}{2})\|Z_k' Z_k\|^{1/2}}{\Gamma(b)\Gamma(c)(2\pi)^{k_j/2}\sigma^{k_j}} \text{ x}$$ $$\Psi(c + \frac{k_j}{2}, \frac{k_j}{2} - b + 1\ ;\ \frac{\beta_k' Z_j' Z_j \beta_j}{2\sigma^2})$$	(2012). Following Fenandez et al. (2001b), they set b/c= max{n, k^2} and they get a Benchmark Beta prior with a fixed value for b=1 and c=1/ max{n,k^2}. This leads to a single hyper-g prior with α = 2* max{n,k^2}+1/ max{n,k^2}=2 + 2/ max{n,k^2} In that case, Beta distribution becomes Beta (b,c)= Beta (1, 1/ max{n,k^2})= Beta (1 , (α-2)/2) It is known as the "hg-BRIC" prior.
23) $E[\frac{g}{1+g}] = \frac{N}{1+N}$ "$hg-UIP$" and $E[\frac{g}{1+g}] = \frac{K^2}{1+K^2}$ "$hg-RIC$"	The shrinkage factor δ = g / (1+g) follows a beta distribution Beta (1, (α/2)-1) The hyperparameter α expresses the prior beliefs on the shrinkage factor δ.	Building on the grounds of Liang et al. (2008), Feldkircher and Zeugner (2009), specify the hyperparameter α as follows: α = 2 + 2/n and α = 2 + 2/k^2 The first one is known as "hg-UIP" and the second one as "hg-RIC".
24)	If the random shrinkage coefficient k_i follows a Beta distribution Beta (b, c)	It is introduced by Carvalho (2010)

$p(k_i) \propto k_i^{-1/2}(1- k_i)^{-1/2}$ $p(\lambda_i) \propto (1 + \lambda_i^2)^{-1}$ under the specification that b=c=1/2	then, the prior for the local shrinkage parameter λ_i is: $p(\lambda_i) \propto \lambda_i^{2c-1}(1 + \lambda_i^2)^{-(b+c)}$	and is known as the "Horseshoe" prior. He specifies the values for b, c = ½ such that the random shrinkage factor k_i ~ Beta (1/2, 1/2)
25) $p(g) = \frac{g^b(1+g)^{-a-z-1}}{B(a+1,z+1)} I_{(0,\infty)}(g)$ with α > -1 , z > -1	The term 1/(1+g) follows a Beta distribution Be (α+1, z+1). Defining new terms such that α+1=b and z+1=c , the term 1/(1+g) follows a Beta distribution Beta (b, c)=Beta (1/4, (n-q-1)/2 – b) where b= α+1= -3/4 +1 = ¼ c= z + 1= (n-5)/2-q/2-α +1 = (n-q-1)/2 – 4/2 – α +1 = (n -q -1)/2 -1 – α= (n -q -1)/2 – (α +1) = (n -q -1)/2 – b and -1 < α < -1/2 → 0 < α +1 < ½→ 0 < b < ½ z > -1 → z+1 >0 → c > 0	It is proposed by Maruyama and George (2011) in the context of formulating selection criteria based on a fully Bayes formulation. For q<n-1, they constrain the choice for α to the interval (-1, -1/2). The default choices for these two hyperparameters are: α= -3/4 and z = (n-5)/2 – q/2 – α.
26) $p(g) = \frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)} g^{b-1}(1+g)^{-(b+c)}$	The shrinkage factor follows a b=Beta distribution δ ~ Be (b, c). The term b=0 and by setting a= 2, the term c= a/2 – 1= 0.	Following Ley and Steel (2012), Bottolo and Richardson (2008) adopt a hyper-g prior with a=2, but make it proper by truncating the right tail at max{n, k^2}.
27) $p(g) = \frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)} g^{b-1}(1+g)^{-(b+c)}$	They recommend the use of b=1 and c=1/2, while truncating g to be $g > \frac{(n+1)}{k_j+3} - 1$	Proposed by Forte, Bayarri, Berger, Garcia-Donato (2010).

Table 2: List of countries

Australia	Algeria
Austria	Brazil
Belgium	B-S-J-G (China)
Canada	Bulgaria
Chile	CABA (Argentina)
Czech Republic	Colombia
Denmark	Costa Rica
Estonia	Croatia
Finland	Cyprus
France	Dominican Republic
Germany	FYROM
Greece	Georgia
Hungary	Hong Kong (China)
Iceland	Indonesia
Ireland	Jordan
Israel	Kosovo
Italy	Lebanon
Japan	Lithuania
Korea	Macao (China)
Latvia	Malta
Luxembourg	Moldova
Mexico	Montenegro
Netherlands	Peru
New Zealand	Qatar
Norway	Romania
Poland	Russia
Portugal	Singapore
Slovak Republic	Chinese Taipei
Slovenia	Thailand
Spain	Trinidad and Tobago
Sweden	Tunisia
Switzerland	United Arab Emirates
Turkey	Uruguay
United Kingdom	Viet Nam
United States	Kazakhstan
Albania	Malaysia

Source: This list of countries is taken from PISA 2015 dataset. More information can be found at http://www.oecd.org/pisa/data/

Table 3: Summary statistics for the dependent ans the independent variables

	Observations	Mean	Std.Dev	Min	Max
Dependent Variable					
Students' performance in science	72	465.297	49.125	331.639	555.575
Independent Variables					
Educational Outcomes					
Repetition of a grade	72	0	1	-1.031	4.055
Non-attendance in pre-primary education	56	0	1	-0.699	5.011
Attendance in pre-primary education 1	56	0	1.755	-3.043	3.569
Attendance in pre-primary education 2	56	0	1.151	-1.897	3.229
Numbers of years in pre-primary education	56	0	1	-2.71	2.036
Participation in Education					
Attendance in schools 1	71	0	1.435	-5.117	1.941
Attendance in schools 2	71	0	1.026	-3.554	2.915
Fields of Education					
Expectations to work in science-related fields 1a	72	0	2.274	-3.435	5.593
Expectations to work in science-related fields 2a	72	0	1.557	-2.756	8.19
Expectations to work in science-related fields 3a	72	0	1.489	-2.96	5.234
Expectations to work in science-related fields 4a	72	0	0.962	-2.764	1.92
Relative risk/Increased likelihood 1b	70	0	1.146	-4.483	2.608
Relative risk/Increased likelihood 2b	70	0	1.049	-4.778	3.885
Relative risk/Increased likelihood 3b	70	0	0.963	-2.764	2.593
Student evaluation and assessment					
Students' evaluation and assessment 1	67	0	1.385	-3.492	2.87
Students' evaluation and assessment 2	67	0	1.374	-2.793	3.867
Students' evaluation and assessment 3	67	0	1.184	-2.486	2.811
Students' evaluation and assessment 4	67	0	0.980	-2.682	2.217
Classroom environment/School climate					
School characteristics 1a	72	0	1.351	-2.115	3.800
School characteristics 2a	72	0	0.849	-2.690	2.730
Average time spent in learning 1b	57	0	1.576	-2.687	6.277
Average time spent in learning 2b	57	0	0.883	-1.865	2.190
Students' engagement, drive and beliefs					
Benefits from science knowledge 1a	57	0	1.698	-4.127	3.66
Benefits from science knowledge 2a	57	0	1.039	-2.613	3.044
Gender difference (boys-girls) in benefits from science knowledge 1b	57	0	1.623	-2.882	3.751
Skip/arrive late in class	57	0	1.795	-4.298	4.092
Skip/arrive late in class	57	0	1.225	-2.282	2.876
After-School activities					

Average time spent in studying after school 1	57	0	1.893	-3.586	5.394
Access to ICT					
Access to ICT	72	0	1	-1.623	2.758
Performance and socio-economic status					
Difference in science performance 1c	61	0	1.549	-3.004	2.768
Difference in science performance 2c	61	0	1.269	-2.716	2.996
Difference in science performance 3c	61	0	1.074	-2.606	2.452
Performance and diversity					
Students with an immigrant background 1a	66	0	1.354	-3.634	3.145
Difference in science performance between immigrant and non-immigrant students 1b	66	0	1.263	-4.093	3.193
Resources for education					
Resources for science course 1a	68	0	1.379	-2.839	2.321
Resources for science course 2a	68	0	1.009	-3.623	1.617
Shortage in resources 1b	72	0	1.415	-3.370	2.647
Shortage in resources 2b	72	0	1.301	-3.084	3.765
Professional Development of Teachers					
Professional Development of Teachers	72	0	1	-1.906	2.122
School evaluation					
Students' evaluation 1a	72	0	1.497	-3.421	3.540
Students' evaluation 2a	72	0	1.102	-1.922	3.442
Governance					
School autonomy	72	0	1	-2.366	1.898
Area of residence as a criterion for admission	72	0	1	-1.664	1.995

Note: The independent variables refer to the new variables that are constructed using the Principal Component Analysis.

Table 4: Three rules for the choice of the number of Principal Components

Rule	Description
Cumulative percentage of total variation	Selection of a cumulative percentage of total variation that the selected PCs contribute, say 60% or 70%.
Size of variances of Principal Components (Kaiser's rule)	Retain only those PCs whose variances exceed one. (the variance for each PC is given by the eigenvalue. Since the values are standardized the average variance/average eigenvalues equal(s) one). Graphically, this is given by the scree plot (ordered eigenvalues). The idea is to keep all eigenvalues

	that form the first portion of the kink (or an elbow) and it is above the red horizontal line.
Eigenvalue Difference graph	Find the average of the differences of successive eigenvalues (it is given by the red horizontal line). Retain all eigenvalues whose differences are above this threshold.

Table 5: Description of the New set of independent variables

Variables	**Notation**	**Number of PC**	**Description**
1) Educational Outcomes			
Repetition of a grade	EducOut1	None	% of students who had repeated a grade in primary, lower secondary or upper secondary education
Non-attendance in pre-primary education	EducOut2	None	% of students who had not attend pre-primary education
Attendance in pre-primary education	PC11, PC21	PC11 is the first principal component (that explains most of the variation) and PC21 is the second in row principal component.	% of students who attend pre-primary education.
Number of years in pre-primary education	EducOut9	None	Number of years that students spend for pre-primary education
2)Participation in Education			
Attendance in schools	PC12, PC22	PC11 is the first principal component (that explains most of the variation) and PC21 is the second in row principal component.	% of students who attend different kind of schools (e.g private, public, government dependent)
3) Fields of Education			
Expectations to work in science-related fields (a)	PC13a, PC23a, PC33a, PC43a	PC13a is the first principal component (that explains most of the variation) and PC23a is the second in row principal component, PC33a is the third and PC43a is the fourth.	% of students who expect to work in science-related professional and technical occupations at the age 30.

Relative risk/increased likelihood (b)	PC13b, PC23b, PC33b	PC13b is the first principal component (that explains most of the variation) and PC23b is the second in row principal component, PC33b is the third.	The relative risk of boys expecting to work in science-related professional and technical occupations at age 30 (expressed in points).
4) Student evaluation and assessment			
Students' evaluation and assessment	PC14, PC24, PC34, PC44	PC14 is the first principal component (that explains most of the variation) and PC24 is the second in row principal component, PC34 is the third and PC44 is the fourth.	The number of students who are evaluated and assessed in alternative ways (e.g. mandatory or no- mandatory tests, teacher-developed tests etc.)
5) Classroom environment/school climate			
School characteristics (a)	PC15a, PC25a	PC15a is the first principal component (that explains most of the variation) and PC25a is the second in row principal component.	Mean ratio defined between students and classroom characteristics.
Average time spent in learning (b)	PC15b, PC25b	PC15b is the first principal component (that explains most of the variation) and PC25b is the second in row principal component.	The average time spent after school in studying science and non-science lessons (expressed in hours).
6) Students' engagement, drive and beliefs			
Benefits from science knowledge (a)	PC16a, PC26a	PC26a is the first principal component (that explains most of the variation) and PC26a is the second in row principal component.	Mean index of students who report that learning about science is beneficial in different ways.
Gender difference (boys-girls) in benefits from science knowledge (b)	PC16b	PC16b is the first principal component (that explains most of the variation).	Mean index of the gender difference (boys-girls) who report that learning about science is beneficial in different ways.

Skip/arrive late in class (c)	PC16c, PC26c	PC16c is the first principal component (that explains most of the variation) and PC26c is the second in row principal component.	The number of students who refer skipping/arriving late at classes two weeks prior to the test.
7) After- School activities			
Average time spent in studying after school	PC17	PC17 is the first principal component (that explains most of the variation).	The average time per week spent after school in studying science and non-science lessons (expressed in hours).
8) Access to ICT			
Access to ICT	ICT	None	Number of computers per student.
9) Performance and socio-economic status			
Difference in science performance (c)	PC19c, PC29c, PC39c	**PC13b is the first principal component (that explains most of the variation) and PC23b is the second in row principal component, PC33b is the third.**	Difference in science performance associated with different reasons. (expressed in score points).
10) Performance and Diversity			
Students with an immigrant background (a)	PC110a	PC110a is the first principal component (that explains most of the variation)	% of students with or without an immigrant background.
Difference in science performance between immigrant and non-immigrant students (b)	PC110b	PC110b is the first principal component (that explains most of the variation).	Difference is science performance between immigrant and non-immigrant students (expressed in score-points).
11) Resources for education			
Resources for science course (a)	PC111a, PC211a	**PC13b is the first principal component (that explains most of the variation) and PC23b is the second in row principal component.**	% of science teachers who are qualified to teach science.
Shortage in resources (b)	PC111b, PC211b	**PC13b is the first principal component (that explains most**	Mean index of shortage in educational material.

		of the variation) and PC23b is the second in row principal component.	
12) Professional Development of teachers			
Professional Development of Teachers	ProfDevTeach	None	% of teachers attended a programme of professional development in the previous three months.
13) School evaluation			
School evaluation	PC113, PC213	PC113 is the first principal component (that explains most of the variation) and PC213 is the second in row principal component.	% of students who use internal/external evaluation.
14) Governance			
School autonomy (a)	Governance1	None	Mean index of school autonomy (% of tasks for which the schools have considerable responsibility)
Area of residence as a criterion of admission	Governance2	None	% of students in schools where residence in a particular area is always considered for admission to school

Table 6: Gini and OLS results using BMA methodology when the dependent variable is Students' performance in Science

	Gini-BMA			OLS- BMA		
	PIP	PSE	PSD	PIP	PSE	PSD
PC110b	0.9389	-19.092	9.224	0.4483	3.878	6.463
PC25a	0.9281	17.136	7.469	0.3110	-0.489	5.543
PC11	0.8167	-8.691	6.976	0.3103	0.053	7.079
PC23a	0.7882	-11.631	8.265	0.2801	-0.672	3.111
PC33b	0.7711	18.040	13.241	0.2663	0.207	3.818
EducOut2	0.7301	-10.446	8.504	0.8377	-8.755	6.666
PC17	0.6860	-5.880	5.476	0.3075	0.101	4.104
PC44	0.6818	5.905	5.530	0.4749	-4.665	6.856
PC110a	0.6486	7.391	7.508	0.9267	-16.699	8.515
PC111a	0.6122	7.398	8.006	0.3201	1.369	3.989
PC25b	0.6069	4.667	5.564	0.3732	2.502	5.526
PC34	0.5621	-4.735	6.151	0.5649	3.872	4.735
EducOut1	0.5381	4.396	5.587	0.7492	-10.257	8.065
PC24	0.4771	-3.226	5.278	0.414	-2.453	5.287
Governance1	0.4764	3.462	5.806	0.4891	-3.121	10.056
PC13a	0.4686	4.233	6.451	0.8904	-14.292	7.359
Governance2	0.4204	-2.621	9.401	0.3410	1.467	3.538
PC16a	0.4199	3.397	6.014	0.3432	-1.966	4.484
PC23b	0.3935	3.955	9.024	0.8636	22.127	12.416
PC15b	0.3910	-2.713	6.989	0.7464	6.324	5.401
PC213	0.3845	1.501	5.685	0.4049	2.319	5.288
PC39c	0.3633	1.373	5.195	0.7080	7.650	6.705
EducOut9	0.3618	1.633	3.545	0.3150	0.304	13.409
PC15a	0.3613	-2.843	5.845	0.9248	16.629	7.451
PC43a	0.3545	-1.375	2.863	0.2363	0.334	2.362
PC113	0.3518	2.692	5.787	0.3321	0.880	5.298
PC19c	0.3511	-2.437	8.052	0.2226	0.289	1.840
ProfDevTeac1	0.3251	0.619	4.800	0.3150	2.142	5.498
PC26a	0.3228	-1.768	4.374	0.2537	0.118	4.206
PC21	0.3123	0.109	7.211	0.3849	1.808	3.607
PC211a	0.3108	1.180	3.954	0.3390	-1.078	3.249
PC12	0.3104	-0.744	13.500	0.2977	0.653	3.107
ICT	0.2931	-0.105	3.936	0.3204	-2.103	6.900
PC211b	0.2816	0.654	2.192	0.3368	0.402	5.050
PC22	0.2802	0.570	2.878	0.3711	2.560	5.558
PC26c	0.2789	0.794	2.922	0.7751	-6.840	5.202

PC111b	0.2785	-0.494	3.038	0.2430	0.268	1.958
PC33a	0.2756	-0.337	2.910	0.4361	-1.935	3.152
PC16c	0.2672	-0.284	2.957	0.2697	0.445	2.956
PC14	0.2668	0.358	3.961	0.354	-1.457	4.319
PC16b	0.2547	-0.403	4.321	0.2571	-0.321	2.790
PC13b	0.2285	0.315	2.246	0.3370	2.352	7.989
PC29c	0.2275	0.149	1.894	0.3556	1.864	4.832

Note: The values in yellow colour are above the 50% PIP and are determined as robust determinants

Table 7: Summary of the robust determinants under Gini and OLS analysis

Variables	Under Gini Analysis	Variables	Under OLS Analysis
Students with an immigrant background	**PC110a (+)**		**PC110a (-)**
Students' evaluation and assessment	**PC34 (-)**		**PC34 (+)**
Repetition of a grade	**EducOut1 (+)**		**EducOut1 (-)**
Non-attendance in pre-primary education	**EducOut2 (-)**		**EducOut2 (-)**
School characteristics	PC25a (+)		PC15a (+)
Average time spent in learning	PC25b (+)		PC15b (+)
Expectations to work in science-related fields	PC23a (-)		PC13a (-)
Relative risk/increased likelihood	PC33b (+)		PC23b (+)
Difference in science performance between immigrant and non-immigrant students	PC110b (-)	Difference in science performance	PC39c (+)
Attendance in pre-primary education	PC11 (-)	Skip/arrive late in class	PC26c (-)
Average time spent in studying after school	PC17(-)		
Students' evaluation and assessment	PC44(+)		
Resources for science course	PC111a (+)		

Notes: There are two variables that refer to pre-primary education and are indicated with yellow color

References:

Aguilar R. and R. Tansini (2012). Joint analysis of preschool attendance and school performance in the short and long-run. International Journal of Educational Development. Vol. 32, pp. 224-231.

Akabayashi H. and R. Tanaka (2013). Long-term effects of pre-schooling on educational attainments. GRIPS Discussion paper, pp. 12-21.

Alexander K. L., D. R. Entwisle, and S. L. Dauber (1994). On the Success of Failure. New York: University of Cambridge Press.

Bainbridge J., M. Meyers, S. Tanaka and J. Waldfogel (2005). Who gets an early education? Family income and the gaps in enrolment of 3-5-year olds from 1968-2000. Social Science Quarterly. Vol. 86(3), pp.724– 745.

Baker M., J. Gruber and K. Milligan (2005). Universal Child Care, Maternal Labour Supply and Family Well-Being. NBER Working Paper No. 11832.

Barnett S. (1992). Benefits of compensatory preschool education. Journal of Human Resources. Vol. 27, pp. 279-312.

Barnett S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. The Future of Children. Vol.5 (3), pp. 25–50.

Barnett W. S. (2011). Effectiveness of early educational intervention. Science. Vol. 333 (6045), pp. 975–977.

Barnett W. S., J.T. Hustedt, K.B. Robin and K.L. Schulman (2004). The state of preschool: 2004 state preschool yearbook. New Brunswick, NJ: NIEER

Bauchmüller R., M. Gørtz and A. W. Rasmussen (2014). Long-run benefits from universal high-quality pre-schooling. Early Childhood Research Quarterly. Vol. 29, pp. 457-470.

Becker G. (1981). A treatise on the Family. Cambridge, Mass.: Harvard University Press.

Becker G. (1985). Human Capital, Effort, and the Sexual Division of Labour. Journal of Labour Economics. Vol. 3 (1, pt.), pp. S33- S58.

Becker G. S. and N. Tomes (1986). Human Capital and the Rise and Fall of Families. Journal of Labour Economics. Vol. 4(3, pt.2), pp. S1-S39.

Begun J. and T.S. Eicher (2008). In search of an environmental Kuznets curve in sulphur dioxide concentrations: a Bayesian model averaging approach. Environment and Development Economics, Vol. 13, pp. 795-822.

Behrman J. R. and N. Birdsall (1983). The quality of schooling: Quantity alone is misleading. The American Economic Review. Vol. 73(5), pp. 928–946.

Berger J.O. and L.R. Pericchi (1996). The intrinsic Bayes Factor for model selection and prediction. Journal of the American Statistical Association. Vol. 91, pp. 109-122.

Berger J.O. and L.R. Pericchi (2001). Objective Bayesian methods for model selection: Introduction and comparison. Model selection. IMS Lecture Notes- Monograph Series. Vol. 38, pp. 135-207.

Berlinski S., S. Galiani and M. Manacorda (2008). Giving children a better start: Preschool attendance and school-age profiles. Journal of Public Economics. Vol. 92, pp. 1416-1440.

Berlinski S., S. Galiani and P. Gertler (2009). The effect of pre-primary education on primary school performance. Journal of Public Economics. Vol. 93, pp. 219-234.

Bernardo J.M. and A.F.M. Smith (1994). Bayesian Theory. John Wiley, Chichester.

Bjorklund A. and K. Salvanes (2011). Education and Family Background: Mechanisms and Policies. in Hanushek, Eric, Machin, Stephen and Ludger Woessmann ed., Handbook of the Economics of Education, Vol. 3, Elsevier.

Blau D. and J. Currie (2006). Pre-school, day care, and after school care: who's minding the kids? In: Hanushek, Eric, Welch, Finis (Eds.), Handbook of the Economics of Education, Elsevier, Amsterdam, Vol.2, pp. 1163–1278.

Bottolo L. and S. Richardson (2008). Fully Bayesian variable selection using g-priors. Working paper. Imperial College.

Brock W.A. and S.N. Durlauf (2001). Growth empirics and reality. The World Bank Economic Review. Vol. 15, pp. 229-272.

Brock W.A, S.N. Durlauf and K.D. West (2003). Policy evaluation in uncertain economic environments. Brookings Papers on Economic Activity, Vol. 2003, pp. 235-301.

Brooks-Gunn J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs? SRCD Social Policy Report, Vol. 17, pp. 3–14.

Brown P.J., M. Vannucci and T.Fearn (1998). Multivariate Bayesian variable selection and prediction. Journal of the Royal Statistical Society. Series B (Statistical Methodology). Vol. 60, pp. 627-641.

Brown P.J., M. Vannucci and T.Fearn (2002). Bayes model averaging with selection of regressors. Journal of the Royal Statistical Society. Series B (Statistical Methodology). Vol. 64, pp. 519-536.

Burhan N.A.S., Y. Kurniawan, A.H. Sidek and M.R. Mohamad (2014). Crimes and the Bell curve: The role of people with high, average, and low intelligence. Intelligence. Vol.47, pp.12-22.

Carcea M. and R. Serfling (2015). A Gini autocovariance function for time series modelling. Journal of Time Series Analysis. Vol. 36, pp. 817-838.

Carneiro P. and J. Heckman (2003). Human Capital Policy. NBER Working Paper No. w9495.

Carvalho C.M., N.G. Polson and J.G. Scott (2010). The horseshoe estimator for sparse signals. Biometrika. Vol. 97, pp. 465-480.

Cascio E. U. (2009). Do investments in universal early education pay off? Long-term effects of introducing kindergartens into public schools. National Bureau of Economic Research, Working Paper 14951.

Ceriani L. and P. Verme (2012). The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. The Journal of Economic Inequality. Vol.10, pp. 421-443.

Charpentier A., N. Ka, S. Mussard and O.H Ndiaye (2019). Gini regressions and Heteroskedasticity. Econometrics. Vol.7, pp.?

Chetty R., J.N. Friedman, N. Hilger, E. Saez, D. Whitmore Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. Quarterly Journal of Economics, Vol. 126(4), pp. 1593–1660.

Chipman H., E.I. George and R.E. McCulloch (2001). The practical implementation of Bayesian Model Selection. In Model Selection (P.Lahiri, ed.). IMS Lecture Notes-Monograph Series. Vol. 38., pp. 70-134.

Clyde M. and E.I. George (2004). Model uncertainty. Statistical Science. Vol. 19, pp. 81-94.

Cui W. and E.I. George (2008). Empirical Bayes vs Fully Bayes variable selection. Journal of Statistical Planning and Inference. Vol. 138, pp. 888-900.

Cumberworth S. and E. Cumberworth (2018). Does school composition matter more for lower-SES Students? A cross- national examination of school socioeconomic composition, individual socioeconomic status, and standardizes test scores.

Cunha F., J. Heckman, L. Lochner and D. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation. in Hanushek, Eric and Finis Welch ed., Handbook of the Economics of Education, Vol. 1, Elsevier.

Currie J. (2001). Early childhood education programs. Journal of Economic Perspectives. Vol. 15, pp. 213–238.

Danziger S. and J. Waldfogel (2000). Investing in children: What do we know? What should we do? Centre for Analysis for Social Exclusion, LSE, CASE papers 34.

Dickson M. (2012). The effect of free pre-school education on children’s subsequent academic performance: Empirical Evidence from England.

Doppelhofer G., R.I. Miller and X. Sala-i-Martin (2004). Determinants of long-term growth: A Bayesian Averaging of Classical Estimates (BACE) approach. American Economic Review. Vol. 94, pp. 813-835.

Drapper D. (1995). Assessment and propagation of model uncertainty. Journal of the Royal Statistical Society. Series B(Methodological). Vol.57, pp.45-97.

Dumas C. and A. Lefranc (2010). Early schooling and later outcomes: Evidence from pre-school extension in France. From parents to children: the intergenerational transmission of advantage, Russel Sage Foundation, 978-0-87154-045-4. ffhal-02528291f

Durlauf S.N., A. Kourtellos and C.M. Tan (2008). Are any growth theories robust?. The Economic Journal. Vol. 118, pp. 329-346.

Durlauf S.N., A. Kourtellos and C.M. Tan (2012). Is God in the details? A Re-approximation of the Role of religion in economic growth. Journal of Applied Econometrics. Vol.27, pp. 1059-1075.

Efroymson M.A. (1960). Multiple Regression Analysis. In Mathematical Methods for Digital Computers, pp. 191-203. Wiley: New York.

Eicher T. S., A. Lenkoski and A.E Raftery (2009). Bayesian model averaging and endogeneity under model uncertainty: an application to development determinants. Working paper No.94, University of Washington.

Eicher T.S., C. Papageorgiou and A.E. Raftery (2011). Default priors and predictive performance in Bayesian Model Averaging, with application to growth determinants. Journal of Applied Econometrics. Vol. 26, pp. 30-35.

Eicher T. S, C. Henn and C. Papageorgiou (2012). Trade Creation and Diversion Revisited: Accounting for model uncertainty and natural trading partner effects. Journal of Applied Econometrics. Vol. 27, pp. 296-321.

Eide E.R. and D. D. Goldhabe (2005). Grade retention: What are the costs and benefits? Journal of Education Finance, Vol. 31 (2), pp. 195-214.

Eklund J. and S. Karlsson (2007). Forecast combination and model averaging using predictive measures. Econometric Reviews. Vol. 26, pp. 329-363.

Entwisle D. and K.L. Alexander (1993). Entry into school: The beginning school transition and educational stratification in the United States. Annual Review of Sociology. Vol. 19, pp.401–423.

Erdogdu F. and E.Edrogdu (2015). The impact of access to ICT, student background and school/home environment on academic success of students in Turkey: an international comparative analysis. Computers & Education. Vol.82, pp. 26-49.

Esping-Andersen G., I. Garfinkel, W.J. Han, K. Magnuson, S. Wagner and J. Waldfogel (2012). Childcare and school performance in Denmark and the United States. Children and Youth Services Review. Vol. 34(3), pp. 576–589.

Feldkircher M. and S. Zeugner (2009). Benchmark priors revisited: On adaptive shrinkage and the supermodel effect on Bayesian Model Averaging. IMF Working paper 09/202.

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

Fernandez C., E. Ley and M.F.J. Steel (2001b). Benchmark priors for Bayesian model averaging. Journal of Econometrics. Vol. 100, pp. 381-427.

Forte A., M.J. Bayarri, J.O. Berger and G. Garcia-Donato (2010). Closed form objective Bayes factors for variable selection in linear models poster presentation. In: Frontiers of Statistical Decision Making and Bayesian Analysis in Honour of Jim Berger.

Carneiro P. and J. Heckman (2003). Human Capital Policy. NBER Working Paper No. w9495.

Garratt A., K. Lee, M.H. Pesaran and Y. Shin (2003). Forecast uncertainties in Macroeconomic modelling: An application to the U.K. economy. Journal of the American Statistical Association. Vol. 98, pp. 829-838.

Garthwaite P.H and e. Mubwandarikwa (2010). Selection of weights for weighted model averaging. Australian & New Zealand Journal of Statistics. Vol. 52, pp. 363-382.

George E.I. and R.E. McCulloch (1993). Variable selection via Gibbs Sampling. Journal of the American Statistical Association. Vol. 88, pp. 881-889.

George E.I. (1999a). Bayesian model selection. Encyclopedia of Statistical Sciences Update. In: Kotz, S., C. Read and D.L. Banks (Eds.), Wiley, New York. Vol. 3, pp. 39-46.

George E.I. (1999b). Discussion of Bayesian model averaging and model search strategies. In: Clyde M.A., J. Bernardo, J. Berger, A. David, and A. Smith (Eds), Bayesian Statistics, Clarendon Press, Oxford. Vol.6, pp.157-177.

George E.I. and D.P. Foster (2000). Calibration and empirical Bayes variable selection. Biometrika. Vol. 87, pp. 731-747.

George E.I. (2010). Dilution priors: Compensating for model space redundancy. Borrowing Strength: Theory Powering Applications- A Festschrift for Lawrence D. Brown. IMS Collections. Vol. 6, pp. 158-165.

Giambona F. and M. Porcu (2018). School size and students' achievement. Empirical evidences from PISA survey data. Socio-Economic Planning Sciences. Vol.64, pp.66-77.

Gilliam W. S. and E. Zigler (2001). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. Early Childhood Research Quarterly. Vol. 15, pp. 441–473.

Goodman A. and B. Sianesi (2005). Early education and children's outcomes: How long do the impacts last? Fiscal Studies. Vol. 26(4), pp. 513-548.

Gormley W.T. and T. Gayer (2005). Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program. Journal of Human Resources. Vol.40(3), pp. 533-558.

Gormley W.T, T. Gayer, D. Phillips and B. Dawson (2005). The effects of universal pre-K on cognitive development. Developmental Psychology. Vol. 41(6), pp. 872–884.

Goux D. and E. Maurin (2008). Preschool enrolment, mother's participation in the labour market and children's subsequent outcomes. mimeo

Grange C.W.J. and H.F. Uhlig (1990). Reasonable Extreme-Bounds Analysis. Journal of Econometrics. Vol. 44, pp. 159-170.

Greene J. P. and M. A. Winters (2004). An Evaluation of Florida's Program to End Social Promotion. Manhattan Institute for Policy Research Education Working Paper No. 7, December.

Hanushek E. and L. Wößmann (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. Economic Journal, Vol. 116, pp. 63-76.

Heckman J. and L. Lochner (2000) Rethinking education and training policy. In: Danziger, S.; Waldfogel, J., editors. Securing the Future. Russell Sage Foundation; New York: 2000.

Heckman J. J., A.B. Krueger and B.M. Friedman (2002). Inequality in America: What role for human capital policies? (14th ed.). Cambridge, MA: MIT Press.

Heckman J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. American Economic Review. Vol. 103 (6), pp. 2052-86.

Helal S., J. Li, L. Liu, E. Ebrahimie, S. Dawson, and D.J. Murray (2019). Identifying key factors of student academic performance by subgroup discovery. International Journal of Data Science and Analytics. Vol. 7, pp. 227-245.

Hettmansperger T.P. (1984). Statistical inference based on ranks. New York: John Wiley and Sons.

Hoeting A., D. Madigan, A.E. Raftery and C. T. Volinsky (1999). Bayesian Model Averaging: A Tutorial. Statistical Science. Vol. 14, pp. 382-401.

Jackman S. and B. Western (1994). Bayesian inference for comparative research. The American Political Science Review. Vol.88, pp. 421-423.

Jacobs B. A. and L. Lefgren (2004). Remedial education and student achievement: A regression-discontinuity analysis. The Review of Economics and Statistics. Vol. 86(1), pp. 226–244.

Jeffreys H. (1961). Theory of Probability, third edition. Clarendon Press, Oxford.

Jerrim J., M. Oliver and S. Sims (2019). The relationship between inquiry-based teaching and students' achievement. New evidence from a longitudinal PISA study in England. Learning and Instruction, Vol. 61, pp. 35-44.

Jolliffe I. T. (2002). Principal Component Analysis. 2nd Edition, Springer, New York.

Karweit N. L. (1999). Grade Retention: Prevalence, Timing, and Effects. CRESPAR Report No. 33. Available online at http://scov.csos.jhu.edu/crespar/reports/report33chapt1.html.

Kass R.E. and A.E. Raftery (1995). Bayes factors. Journal of the American Statistical Association. Vol. 90, pp. 773-795.

Kass R.E. and L. Wasserman (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. Journal of the American Statistical Association. Vol. 90, pp. 928-934.

Knight A. and D. Hughes (1995). Developing social competence in the pre-school years. Australian Journal of Early Childhood. Vol. 20 (2), pp. 13–19.

Knudsen E. I., J.J. Heckman, J.L. Cameron and J.P. Shonkoff (2006). Economic, neuro-biological and behavioural perspectives on building America's future workforce. Proceedings of the National Academy of Sciences. Vol. 103(27), pp. 10155–10162.

Koop G. and D. Korobolis (2012). Forecasting inflation using Dynamic Model Averaging. International Economic Review. Vol. 53, pp. 867-886.

Kourtellos A., T. Stengos and C.M Tan (2013). The effect of Public Debt in Growth in Multiple Regimes. Journal of Macroeconomics. Vol.38 (Part A), pp.35-43.

Kourtellos A., C. Marr and C.M. Tan (2015). Robust determinants of intergenerational mobility in the land of opportunity. European Economic Review. Vol. 81, pp. 132-147.

Kourtellos A. and C.G. Tsangarides (2017). Robust correlates of growth spells: Do inequality and redistribution matter?. The Rimini Centre for Economic Analysis. Working paper 15-20.

Laud P.W. and J.G. Ibrahim (1996). Predictive specification of prior model probabilities in variable selection. Biometrika. Vol. 83, pp. 267-274.

Leamer E.E. (1978). Specification searches: Ad hoc inference with nonexperimental data. New York: John Wiley and Sons.

Leamer E.E. (1983). Let's take the con out of Econometrics. The American Economic Review. Vol.73, pp. 31-43.

Leamer E.E. (1985). Sensitivity analyses would help. The American Economic Review. Vol. 75, pp. 308-313.

Lerman R.I and S. Yitzhaki (1984). A note on the calculation and interpretation of the Gini index. Economics Letters. Vol. 15, pp. 363-368.

Levine R. and D. Renelt (1992). A sensitivity analysis of cross-country growth regressions. The American Economic Review. Vol. 82, pp. 942-963.

Levin A.T. and J.C. Williams (2003). Robust monetary policy with competing reference models. Journal of Monetary Economics. Vol. 50, pp. 945-975.

Ley E. and M.F.J. Steel (2007). Jointness in Bayesian variable selection with applications to growth regression. Journal of Macroeconomics. Vol. 29, pp. 476-493.

Ley E. and M.F.J. Steel (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. Journal of Applied Econometrics. Vol. 24, pp. 651-674.

Ley E. and M.F.J. Steel (2012). Mixtures of g-priors for Bayesian model averaging with economic applications. Journal of Econometrics. Vol. 171, pp. 251-266.

Liang F., R. Paulo, G. Molina, M.A. Clyde and J.O. Berger (2008). Mixtures of g priors for Bayesian variable selection. Journal of the American Statistical Association. Vol. 103, pp. 410-423.

Lorence J., A.G. Dworkin, L.A. Toenjes and A.N. Hill (2002). Grade retention and social promotion in Texas, 1994–1999: Academic achievement among elementary school students. In D. Ravitch (Ed.), Brookings papers on education policy 2002 (pp. 13–52). Washington, DC: Brookings Institution.

Lorence J. and A.G. Dworkin (2006). Elementary Grade Retention in Texas and Reading Achievement Among Racial Groups: 1994–2002. Review of Policy Research. Vol.23 (5), pp. 999-1033.

Ludwig J. and D.L. Miller (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. Quarterly Journal of Economics. Vol. 122, pp. 159–208.

Madigan D. and A.E. Raftery (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. Journal of the American Statistical Association. Vol. 89, pp. 1535-1546.

Magnuson K. and J. Waldfogel (2005). Childcare, early education, and racial/ethnic test score gaps at the beginning of school. The Future of Children. Vol. 15(1), pp.169–196.

Magnuson K. A., C. Ruhm and J. Waldfogel (2007). Does prekindergarten improve school preparation and performance? Economics of Education Review. Vol. 26(1), pp. 33 – 51.

Maruyama Y. and E.I. George (2011). Fully Bayes factors with a generalized g-prior. The Annals of Statistics. Vol. 39, pp. 2740-2765.

Massanjala W.H. and C. Papageorgiou (2008). Rough and lonely road to prosperity: A re-examination of the sources of growth in Africa using Bayesian Model Averaging. Journal of Applied Econometrics. Vol. 23, pp. 671-682.

Meyers M., D. Rosenbaum, C. Ruhm and J. Waldfogel (2004). Inequality in early childhood education and care: What do we know? In: Neckerman, K., editor. Social Inequality. Russell Sage Foundation, New York: 2004.

Min C. and A. Zellner (1993). Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. Journal of Econometrics. Vol. 56, pp. 89-118.

Mitchell T.J. and J.J. Beauchamp (1988). Bayesian variable selection in linear regression. Journal of the American Statistical Association. Vol. 83, pp. 1023-1032.

Moser M. and P. Hofmarcher (2014). Model priors revisited: Interaction terms in BMA growth applications. Journal of Applied Econometrics. Vol. 29, pp. 344-347.

Mussard S. and O.H. Ndiaye (2018). Vector autoregressive models: A Gini approach. Physica A: Statistical Mechanics and its Applications. Vol. 492, pp. 1967-1979.

Myers R. (1992). The Twelve who Survive: Strengthening Programmes of Early Childhood Development in the Third World. Routledge, London.

Myers R. (1995). Preschool Education in Latin America: Estate of Practice. PREAL Working Papers No. 1.

Neuman M., S.B. Kamerman, J. Waldfogel and J. Brooks-Gunn (2003). Social policies, family types, and child outcomes in selected OECD countries. OECD Social, Employment, and Migration Working Paper No. 6.

Nikolaev B. and R. Salahodjaev (2016). The role of intelligence in the distribution of national happiness. Intelligence. Vol. 56, pp. 38-45.

OECD (2009). PISA Data Analysis Manual: SPSS Second Edition. OECD Publishing

O'Hagan A. (1995). Fractional Bayes Factors for model comparison. Journal of the Royal Statistical Society. Series B(Methodological). Vol. 57, pp. 99-138.

Olkin I. and S. Yitzhaki (1992). Gini Regression Analysis. International Statistical Review. Vol. 60, pp.185-196.

Pholphirul P. (2017). Pre-primary education and long-term education performance: Evidence from Programme for International Student Assessment (PISA) Thailand. Journal of Early Childhood Research. Vol. 15 (4), pp. 410-432.

Raftery A.E. (1988). Approximate Bayes factors for Generalized Linear Models. Technical Report 121, University of Washington, Department of Statistics.

Raftery A.E. (1993). Bayesian model selection in structural equation models. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp.163-180). Newbury Park, CA: Sage.

Raftery A.E. (1995). Bayesian model selection in social research. Sociological Methodology. Vol. 25, pp. 111-163.

Raftery A.E, D. Madigan and J.A. Hoeting (1997). Bayesian model averaging for linear regression models. Journal of the American Statistical Association. Vol. 92, pp. 179-191.

Reynolds A. J. (1993). Effects of a preschool plus follow-on intervention program for children at risk. Developmental Psychology. Vol. 30, pp. 787–804.

Reynolds A. J. (2000). Child, youth, and family services. Success in early intervention: The Chicago child-parent centers. University of Nebraska Press.

Reynolds A. J. (2000). Success in early intervention: The Chicago child–parent centers. Lincoln. NE: University of Nebraska Press.

Reynolds A.J, J.A. Temple, S.R. Ou, I.A. Arteaga and B.A.B White (2011). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. Science, Vol. 333(6040), pp. 360-364.

Roberts H.V. (1965). Probabilistic prediction. Journal of the American Statistical Association. Vol. 60, pp.50-62.

Sala-i-Martin X.X. (1997). I just ran two million regressions. The American Economic Review. Vol. 87, pp.178-183.

Schulman K., H. Blank and D. Ewen (1999). Seeds of Success: State Prekindergarten Initiatives 1998– 1999. Washington, DC: Children's Defense Fund.

Serfling R. (2010). Fitting autoregressive models via Yule-Walker equations allowing heavy tail innovations. Working paper.

Shelef A. and E. Schechtman (2011). A Gini-based methodology for identifying and analysing time series with non-normal innovations. SSRN Electronic Journal. Vol.?, pp. 1-26.

Sirimaneetham V. and J. Temple (2006). Macroeconomic policy and the distribution of growth rates. CEPR Discussion paper 5642.

Sholderer O. (2017). Making education work: School autonomy and performance. East European Quarterly. Vol.45, pp. 27-56.

Strawderman W.E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. The Annals of Mathematical Statistics. Vol. 42, pp. 385-388.

Stuart A. and J.K Ord (1987). Kendall's advanced theory of statistics, Vol.1, 5th ed., New York: Oxford University Press.

Taiwo A.A and J.B. Tyolo (2002). The effect of pre-school education on academic performance in primary school: a case study of grade one pupils in Botswana. International Journal of Educational Development. Vol. 22, pp. 169-180.

Tang Y. (2019). Immigration Status and Adolescent Life Satisfaction: An International Comparative Analysis Based on PISA 2015. Journal of Happiness Studies. Vol.20, pp. 1499-1518.

Tinajero C., S.M. Lemos, M. Araújo, M.J. Ferraces and M.F. Páramo (2012). Cognitive style and learning strategies as factors which affect academic achievement of Brazilian university students. Psicologia: Reflexão e Crítica, Vol.25, pp. 105–113.

Tobias J.L. and M. Li (2004). Returns to schooling and Bayesian Model Averaging: A union of two literatures. Journal of Economic Surveys. Vol. 18, pp. 153-180.

UNESCO (2005). EFA Global Monitoring Report. (Data available at: http://portal.unesco.org).

Velez E., E. Schiefelbein, J. Valenzuela (1993). Factors Affecting Achievement in Primary Education. World Bank, Washington, Working Paper No. 12186.

Volinsky C.T., D. Madigan, A.E. Raftery and R. A. Kronmal (1997). Bayesian Model Averaging in proportional hazard models: Assessing the risk of a stroke. Applied Statistics. Vol. 46, pp. 433-448.

Waldfogel J. (2002). Childcare, women's employment and child outcomes. Journal of Population Economics. Vol. 15, pp. 527–548.

Waldfogel J. (2006). What Children Need. Harvard University Press, Cambridge, Mass.

Waldfogel J. and E. Washbrook (2011). Income-related gaps in school readiness in the US and UK. In: Smeeding, T., et al., editors. Intergenerational Mobility Within and Across Nations. Russell Sage Foundation; New York: In press b

Wößmann L. (2005). Families, School, and Primary-School Learning: Evidence for Argentina and Colombia in an International Perspective. The World Bank, Policy Research Working Paper 3537.

Yang K.E and S.H. Ham (2017). Truancy as systemic discrimination: Anti-discrimination legislation and its effect on school attendance among immigrant children. The Social Science Journal. Vol. 54, pp.216-226.

Yitzhaki S. and E. Schechtman (2004). The Gini instrumental variable, or the "double instrumental variable" estimator. Metron- International Journal of Statistics. Vol. LXII, pp. 287-313.

Yitzhaki S. and E. Schechtman (2013). The Gini Methodology: a primer on a Statistical Methodology. New York: Springer.

Zellner A. (1971). An introduction to Bayesian Inference in Econometrics. Wiley, New York.

Zellner A. and A. Siow (1980). Posterior odds ratios for selected regression hypotheses (with discussion). In: Bernardo J.M., M.H. DeGroot, D.V. Lindley and A.F.M. Smith (Eds.), Bayesian Statistics. University Press, Valencia, pp. 585-603.

Zellner A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: Goel P.K. and A. Zellner (Eds.), Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti. North-Holland, Amsterdam, pp. 233-243.

Zheng A., E.M Tucker-Drob and D.A Briley (2019). National GDP, Science Interest, and Science Achievement: A Direct Replication and Extension of Tucker-Drob, Cheung and Briley (2014). Psychological Science, Vol. 30, pp. 776-788.

Zigler E. and W. Berman (1983). Discerning the future of early childhood intervention. American Psychologist. Vol. 38, pp. 894-906.