MANKIND QUARTERLY 2022 63.1 9-78

National Intelligence and Economic Growth: A Bayesian Update

George Francis*
Independent Researcher, United Kingdom

Emil O.W. Kirkegaard*
Independent Researcher, Denmark

* Addresses for contact: george.t.francis@protonmail.com and emil@emilkirkegaard.dk

Since Lynn and Vanhanen's book *IQ and the Wealth of Nations* (2002), many publications have evidenced a relationship between national IQ and national prosperity. The strongest statistical case for this lies in Jones and Schneider's (2006) use of Bayesian model averaging to run thousands of regressions on GDP growth (1960-1996), using different combinations of explanatory variables. This generated a weighted average over many regressions to create estimates robust to the problem of model uncertainty. We replicate and extend Jones and Schneider's work with many new robustness tests, including new variables, different time periods, different priors and different estimates of average national intelligence. We find national IQ to be the "best predictor" of economic growth, with a higher average coefficient and average posterior inclusion probability than all other tested variables (over 67) in every test run. Our best estimates find a one point increase in IQ is associated with a 7.8% increase in GDP per capita, above Jones and Schneider's estimate of 6.1%. We tested the causality of national IQs using three different instrumental variables: cranial capacity, ancestry-adjusted UV radiation, and 19th-century numeracy scores. We found little evidence for reverse causation, with only ancestry-adjusted UV radiation passing the Wu-Hausman test ($p < .05$) when the logarithm of GDP per capita in 1960 was used as the only control variable.

Key Words: Human capital, National IQ, Economic growth, Bayesian model averaging, Intelligence, Smart fraction

Much research has found that national IQ is associated with per capita GDP and economic growth, suggesting cognitive skills are important for prosperity. However, it is unclear exactly how robust this relationship is and how important this theory is compared to alternative theories. Merely reporting regression models is not sufficient. Researchers can choose what variables to employ and what models to report that best support their own theories. Moreover, only so many variables can be included in a model, allowing for few comparisons between national IQ and other explanatory variables.

Jones and Schneider (2006) came closest to addressing this problem. They performed Bayesian model averaging of economic growth, running thousands of regressions with different explanatory variables, including national IQ, and then took a weighted average of these results. National IQ was in 96% of the models after the models were weighted according to how well they fitted the data, suggesting national IQ is very good at predicting economic growth. Unfortunately, Jones and Schneider did not report results for other variables making any firm comparisons between national IQ and other variables difficult. Furthermore, recent literature on Bayesian model averaging and economic growth has found the results to be very sensitive to minor changes in the data and priors used (Bruns & Ioannidis, 2020; Ciccone & Jarociński, 2010; Rockey & Temple, 2016). Just as with regression models, it is not sufficient to only present one or a few Bayesian model averages, rather a large range should be used if we are to be confident in the results.

In this paper, we replicate and extend Jones and Schneider's (2006) work by employing national IQ in Bayesian model averaging. We standardize our variables and present the results for all variables possible to compare national IQ to rival theories of economic growth. We use a long series of stress tests from the Bayesian model averaging literature and others we have come up with. This includes resampling, different datasets, different priors, and different time periods. This is to see if national IQ is robust or whether Jones and Schneider's results were merely a fluke from the sensitivity of Bayesian model averaging.

We also review the prior literature on causality and use instrumental variables to test this. GDP per capita at the start of the observation period is used as a "fixed regressor" in all regressions, meaning it is not a variable we test in this paper. We use the term fixed regressor not in the traditional way it is used in time series econometrics, but rather to mean that it is forced into every regression that is model averaged, ensuring its posterior inclusion probability (PIP) is always equal to one.

Compared to all other tested explanatory variables, we find national IQ to be the most important predictor of economic growth. In every set of stress tests, we find national IQ to have the largest average coefficient and the largest posterior inclusion probability. Our results suggest each IQ point increases GDP per capita by 7.8% compared to Jones and Schneider's 6.1% estimate. We extend the use of Bayesian model averaging to re-evaluate many questions in the national IQ literature. We test smart fraction theory — whether the mean IQ or that of another part of the IQ distribution best predicts economic growth, as well as rival psychological explanations of economic growth, finding national IQ to dominate them all. We also test whether national IQ affects economic growth through an exogenous model, an endogenous one, or a Nelson-Phelps technology diffusion model.

In section 1 we summarize the literature on how national IQ can function as a measure of human capital. In section 2 we review the issue of model uncertainty and how Bayesian model averaging has tried to respond to this problem. In section 3 we explain our Bayesian model averaging methodology. In section 4 we explain what data we use in this paper. This includes which national IQ measures are used, such as the World Bank's Harmonised Learning Outcomes, what measures of GDP we use, what datasets of control variables we employ, and the additional variables we employ in stress tests. In section 5 we present our main results using Bayesian model averaging to test the robustness of national IQ and Smart Fraction theory. Section 6 discusses the problem of causality and tests this with instrumental variables. In section 7 we review the general limitations of our methodology. In section 8 we draw our conclusions and describe their implications for policy and the future of the world economy. In an online supplement we discuss our results in the context of endogenous and exogenous models and test the Nelson-Phelps technology diffusion model.

Section 1: National IQ as a Measure of Human Capital

Economists have empirically tested the role of human capital in causing economic growth since at least Mankiw, Romer and Weil (1992). Typically studies have used years in education, such as Barro and Lee's years in education measure (1993), or enrollment rates in education as in Sala-i-Martin (1997) and Sala-i-Martin, Doppelhofer and Miller (2004). However, despite the correlation between the level of education and economic growth, increases in education have sometimes been found to have no relationship with economic growth (Hamilton & Monteagudo, 1998) or even a negative relationship (Pritchett, 2001), calling into question whether variation in education is a satisfactory indicator of variation in human capital.

An alternative approach to measuring human capital as a predictor of economic growth has been to use measures of human intelligence or attained academic ability rather than amount of schooling — focusing on educational 'output' (Hanushek & Woessmann, 2015), rather than any supposed 'inputs'. If education is ineffective at improving ability (see Caplan, 2018 for a review), or is heterogeneous in its quality, or simply only one of many causes of variation in human ability, then a direct measurement of a people's cognitive ability may better measure their human capital. If more capable countries tend to have more education, then it may be confounded with human ability explaining education's apparently spurious relationship with growth.

Although this output-based approach to human capital had been suggested by Barbara Lerner (1983), its relationship with economic output was only tested and supported in regression models by Hanushek and Kim (1995) and Lynn and Vanhanen (2002). Lynn and Vanhanen created "national IQs" using samples of psychometric IQ tests from countries around the world to create average national IQs to test their effect on GDP. By contrast, Hanushek and Kim used the results of recurrent international student assessments in mathematics and reading to create measures of education quality to predict economic growth.

Many further studies have been published studying economic growth with data from student assessments such as PIRLS, TIMMS and the OECD's Programme for International Student Assessment (PISA) (e.g. Angrist et al., 2019, 2021; Hanushek & Woessmann, 2015; Lim et al., 2018) and IQ tests (e.g. Christainsen, 2020; Meisenberg, 2012; Ram, 2007, Weede & Kämpf, 2002), finding these variables to have positive and significant coefficients. Furthermore, some studies have used both student assessments and national IQs or even combined them (Rindermann, 2008, 2018). Both student assessments and national IQs appear to measure the same construct, cognitive ability (Rindermann, 2006, 2007), as they correlate at 0.8 or more at the national level (Meisenberg & Lynn, 2011). For the sake of ease, we refer to all measures of national average cognitive ability as 'national IQ', rather than just average IQ scores from nationally representative samples.[1]

There are strong theoretical reasons for supposing that IQ would make an appropriate measure of human capital. For a start, intelligent people tend to be more productive workers. Psychologists have found smarter workers are more productive in their occupations (Gottfredson, 1986, 1997). One IQ point is associated with approximately 1% higher wages (Behrman et al., 2004; Bishop,

[1] IQ stands for intelligence quotient and was originally invented to measure cognitive ability by dividing scores by the test taker's age, hence the term 'quotient'. The term is somewhat of a misnomer when used in 'national IQ' because modern IQ tests do not divide by the test taker's age and they are designed for individuals rather than nations. Nevertheless, it has become commonplace to refer to national cognitive ability as national IQ, so we follow this practice.

1989; Cawley et al., 1997; Grosse et al., 2002; Neal & Johnson, 1996; Zax & Rees, 2002).

It is clear that individual differences in intelligence could cause large individual differences in income. However, the magnitude of differences in the average intelligence of nations also implies large differences of wealth between nations. For example, in Lynn's original IQ dataset (Lynn & Vanhanen, 2002), the UK has a national IQ of 100 whilst Brazil has an IQ of 82, more than a standard deviation (15 points) lower than the UK.

TEST QUESTIONS (LEVEL 3)
WHICH CAR?

Chris has just received her car driving licence and wants to buy her first car.

This table below shows the details of four cars she finds at a local car dealer.

Model:	Alpha	Bolte	Castel	Dezal
Year	2003	2000	2001	1999
Advertised price (zeds)	4800	4450	4250	3990
Distance travelled (kilometres)	105 000	115 000	128 000	109 000
Engine capacity (litres)	1.79	1.796	1.82	1.783

QUESTION

Which car's engine capacity is the smallest?

A Alpha

B Bolte

C Castel

D Dezal

Figure 1. *A 'level 3' difficulty question asked on PISA.*

The magnitude and importance of national IQ differences may be more clearly intuited by looking at the national results from individual questions on intelligence tests. Take the question in Figure 1 from the OECD's 2012 maths PISA test given to 15-year-olds (accessed from: https://www.oecd.org/pisa/test-2012/). It involves reading a table providing details regarding different cars. The test subject had to find the car with the smallest engine capacity. This question is considered to be of ‘level 3 difficulty’ by the OECD. Despite the ease of this task, only 55% of OECD students scored level 3 or above. 80% of Singaporeans achieved level 3 or above compared to only 18% of Mexicans, 13% of Brazilians, and 8% of Indonesians. Inability to read tables would make any sort of analytical work impossible and would even make simple tasks, such as reading a train timetable, difficult for the majority of people in many countries. Given the extraordinary differences in average cognitive ability between countries, we ought to consider whether it could affect gross domestic product.

IQ’s effect on individual productivity is not the only explanation for why it predicts economic growth. Jones and Schneider (2010) calibrated an IQ-augmented Cobb-Douglas production function to model log GDP per worker in the year 2000, using the 1% estimate of IQ’s effect on earnings. They found this estimate of IQ’s effect could only explain 16% of the variation in national earnings, whilst in a simple correlation (Lynn & Vanhanen, 2006) IQ can explain 58% of the variation and each additional national IQ point is associated with 6.7% higher per capita earnings. This “IQ paradox” suggests substantial reverse causation or externalities, with intelligent people not fully internalizing the benefits of their intelligence. We discuss the issue of causality in the ‘Causality’ section of this paper.

Potential causes of externalities arising from IQ include its association with free-market opinions (Carl, 2014a, 2015), with greater knowledge of economics (Caplan & Miller, 2010), with lower time preference and more saving (Jones & Podemska-Midluch, 2010; Kirkegaard & Karlin, 2020; Shamosh & Gray, 2008; Yeh et al., 2021), with higher levels of social trust (Carl, 2014b; Carl & Billari, 2014), with cooperation in public goods games (Al-Ubaydli et al., 2016; Putterman et al., 2012) and the prisoner's dilemma (Proto et al., 2019), national IQ’s association with institutional quality (Jones & Potrafke, 2014; Kanyama, 2014; Potrafke, 2012) and the prevalence of O-ring production functions (Jones, 2013). For an overview of the mechanisms by which IQ could create externalities see Jones (2016) and Anomaly and Jones (2021).

Section 2: Modelling Uncertainty

In growth modelling with IQ, and regression modelling generally, researchers are faced with the problem of model uncertainty. Model estimates are dependent on what variables are included, meaning reported significant results from a subset of all possible models may just be an artefact. For example, with 50 potential explanatory variables, there are 2^{50} possible models, which is greater than one quadrillion. How can we be sure the few models presented in a paper, or even an entire literature, are representative and that the results are not just the result of data mining?

Model uncertainty implies that regression results could be coincidental, and it also allows researchers to 'p-hack' their results with specification searching — only presenting the models which best support their theories. In addition, journals may prefer only to publish statistically significant results rather than null findings. The economics literature seems to have been affected by specification searching, as p and t values in published findings are less likely to be marginally insignificant than would be expected by chance (Brodeur et al., 2016; Bruns et al., 2019; Vivalt, 2019). This indicates that economists are biased, consciously or not, and slightly change their methodology until their results are statistically significant. This should make us skeptical of the reported results from merely a few models.

Bayesian model averaging can potentially provide more accurate results free from biases arising from selective model reporting (Zeugner & Feldkircher, 2015). This methodology was first employed to study economic growth by Fernandez et al. (2001) and Sala-i-Martin et al. (2004), although previous papers had attempted similar methods by creating summary statistics from running many (Levine & Renelt, 1992) or millions of regressions (Sala-i-Martin, 1997a,b).

This method involves running a large sample of possible models with different explanatory variables and fixing a prior probability of each model being the 'true model' before looking at the data. Then using Bayes theorem, a posterior model probability is calculated for each model based on their marginal likelihoods. The coefficients of explanatory variables are then calculated by weighting coefficients in individual models by the model's posterior probability. Furthermore, a posterior inclusion probability (PIP) is calculated for each explanatory variable which represents the sum of the posterior model probabilities of the models in which the covariate is included. As such the PIP represents the explanatory power of a variable and can be crudely understood as the probability of a variable being in the 'true model' or the probability that a variable's true coefficient is non-zero. Moreover, by comparing the posterior inclusion probability of a variable to its prior inclusion probability we can see whether the data tends to increase or decrease our impression of whether a variable is a robust predictor. Full details of Bayesian model averaging are given in the Methodology section.

To test how robust national IQ is as a predictor of economic growth, Jones and Schneider (2006) used this variable in Bayesian model averaging of the growth rate in GDP per capita. In addition to national IQ, they employed the 21 variables considered robust in Sala-i-Martin's (1997a,b) paper 'I just ran two million regressions'. Jones and Schneider found that national IQ had an extremely high posterior-inclusion probability of 96%. By comparison, the top three performing variables in Fernandez and Lay's (2001) study using Sala-i-Martin's entire dataset had PIPs of 99.5% (fraction Confucian), 99.5% (life expectancy) and 94% (equipment investment). A 1 point increase in national IQ was associated with a 0.11% increase in the GDP growth rate.

The results of Jones and Schneider (2006) provide strong support for the relationship between economic growth and national IQ and suggest the significant results of other studies are not the result of poorly specified models, whether by coincidence or various forms of bias. Unfortunately, Jones and Schneider do not provide the results for all the control variables, meaning that fifteen years on we still do not have a clear picture of how important national IQ is relative to other variables — both in the size of its coefficient and its posterior inclusion probability. Does IQ have the largest effect size? Are there many or any variables as robust as national IQ? We hope to answer these questions.

The Bayesian modelling literature

Since Fernandez et al. (2001), Bayesian model averaging has been widely applied to the study of economic growth, with focus on things such as the jointness of growth determinants (Doppelhofer & Weeks, 2009; Ley & Steel, 2007), growth in specific regions (Cuaresma et al., 2009, 2013; Masanjala & Papageorgiou, 2008; Próchniak & Witkowski, 2013a) and testing particular theories of economic growth (Durlauf et al., 2008; Égert, 2015; Eris & Ulasan, 2013; Horvarth, 2011; Próchniak & Witkowski, 2013b). However, despite the strong results of Jones and Schneider (2006), no one has employed national IQ with Bayesian model averaging since.

Despite the popular use of Bayesian modelling, some researchers have found the results to be extremely fragile to minor differences in methodology and data. If the results of Bayesian modelling are unstable or unreliable, we may not have confidence that Jones and Schneider's (2006) results on IQ will replicate. Ciccone and Jarocinski (2010) in their paper *Determinants of Economic Growth: Will Data Tell?* re-ran the Bayesian modelling of Sala-i-Martin et al. (2004) using different updates of the Penn World Tables (PWT versions 6.0, 6.1 and 6.2), which provides the national accounting data for variables such as GDP. They found that minor changes in the national accounting data were enough to substantially alter the results of Bayesian model averaging. An extreme example of this was the 'Investment Price' variable, which moved from having a posterior inclusion probability of 98% to 2% depending upon the PWT dataset used. Thus variables that could appear to be robust predictors of economic growth may not be robust to measurement error.

Bruns and Ioannidis (2020) tested the sensitivity of Bayesian modelling using multiple time periods from 1960 to 2010. They found the inferences on growth determinants were unstable across time periods. The posterior inclusion probabilities of determinants were evenly distributed across early time periods preventing identification of which control variables were important. However, Ioannidis and Bruns find more recent time periods show less even distributions of posterior inclusion probabilities suggesting improvements in measurement could cause Bayesian modeling to find more stable results.

In response to concerns about the sensitivity of Bayesian model averaging some researchers have suggested methods to improve BMA. Feldkircher and Zeugner (2009) proposed the use of flexible priors on model coefficients that allow for 'data-adaptive shrinkage'. Feldkircher and Zeugner (2012) show that the use of these priors makes Bayesian modelling more robust to changes in the Penn World Tables. Likewise, Rockey and Temple (2016) recommend the use of fixed regressors to improve the stability of results, particularly the use of GDP per capita in the starting year and regional dummies. However the use of flexible priors and fixed regressors leads to PIPs being more evenly distributed making it difficult to determine which variables are important.

In overview, Bayesian model averaging appears to be either very sensitive, or it fails to discriminate between variables (Bruns & Ioannidis, 2020; Rockey & Temple, 2016), resulting in 'robust ambiguity' with failure to find support for growth regressors, except for GDP per capita in the starting year. Given that national IQ has only been tested with Bayesian modelling by Jones and Schneider (2006), it is reasonable to question whether it will replicate given the apparent conclusion of robust ambiguity. To see whether national IQ's posterior inclusion probability is too sensitive to be robust, we run many sets of Bayesian model averaging, using different datasets, different time periods, different versions of the Penn World Tables, different priors, different fixed regressors, new potential confounds of IQ, different national IQs, and resampling methods such as bootstrapping and Jackknife resampling. In applying all the robustness tests in the BMA growth literature and more, we ensure that our research employs the most challenging set of tests in the literature so far. If national IQ performs the best in all these tests, with the highest average PIP and coefficient in each test, we will consider it to be the most powerful predictor of economic growth in the literature.

Ciccone and Jarocinski (2010) asked in the title of their paper *Determinants of Economic Growth: Will Data Tell?.* We question whether the right data, national IQ, is being used.

Section 3: Methodology

To perform Bayesian modelling averaging (BMA), we use the *BMS* package in R. We draw heavily from the package's tutorial (Zeugner & Feldkircher, 2015) in explaining the methodology.

To model growth we can use a linear regression model of the following structure:

$$y_i = \alpha + X_j\beta_j + \varepsilon_i$$
$$\epsilon \sim N(0, \sigma^2 I)$$
$$X_j \in \{X\}$$

y_i is our dependent variable, the average rate of growth in nations. α is a constant, β_j the regression coefficients, and ε is the normal IID error term[2] of variance σ^2. With many plausible control variables $\{X\}$ to employ as X_j but with a limited sample size, simply employing all variables would not be feasible.

BMA deals with the problem of model uncertainty by estimating models for many possible combinations of $X_j \in \{X\}$ and then taking a weighted average over all of them. The models are weighted by their posterior model probability $P(M_j|y)$ which can be derived from Bayes' Rule:

$$P(M_j|y) = \frac{P(y|M_j, X)P(M_j)}{P(y|X)} = \frac{P(y|M_j, X)P(M_j)}{\sum_{j=1}^{2^k} P(y|M_j, X)P(M_j)}$$

$P(M_j)$ is the model prior and $P(y|M_j)$ is the marginal likelihood of model M_j. $P(y|X)$ denotes the integrated likelihood which is the same for all models. The posterior model probability is thus proportional to the marginal likelihood of the model times the prior model probability.

The marginal likelihood of model M_j is given by

$$P(y|M_j) = \int P(y|\alpha, \beta_j, \sigma, M_j)P(\alpha, \beta_j, \sigma|M_j)d\alpha d\beta_j d\sigma$$

where $P(y|\alpha, \beta_j, \sigma, M_j)$ is the likelihood of model M_j and $P(\alpha, \beta_j, \sigma|M_j)$ is the prior density of the coefficients of model M_j.

The key results we are interested in are the variable posterior inclusion probabilities (PIPs) and the variable coefficients. The PIPs indicate what percent of the posterior model mass is made up of models including regressor X_k.

---

2 IID normality of error terms is certainly a strong assumption, nevertheless there has been little discussion of its importance in the growth literature and the statistical package we employ relies on this assumption. As such, we follow common practice by keeping the assumption.

To be precise the PIP of X_k is the sum of posterior model probabilities over those models where X_k is included.

$$P(I_k = 1|y) = \sum_A P(M_j|y)$$
$$A = \{M_j : j = 1, K, 2^k; I_k\}$$

I_k is an indicator function that is 1 if regressor X_k is included in model j and 0 otherwise. A is the set of models that include X_k.

How should we interpret PIPs? Sala-i-Martin et al. (2004) suggest comparing the posterior inclusion probability to the prior inclusion probability. If the PIP is larger than the prior inclusion probability, we can say that the data has updated our priors in favor of variable X_k. Sala-i-Martin suggests we use this as a threshold for 'significance'.

$$P(\beta_k|y, X) = \sum_{j=1}^{2^k} P(\beta_k|M_j, y, X)P(M_j|X, y)$$

The model weighted density function for coefficients is represented above. $P(\beta_k|M_j, y, X)$ represents the density function of coefficient β_k given the data in model M_j, whilst $P(M_j|X, y)$ represents the marginal likelihood of model M_j.

This explains the core mechanics of Bayesian modeling, but we need to elicit what the model priors are to calculate the results. To perform BMA well we should choose priors that are non-informative so that the priors have little impact on posterior inference so we let the data come to its own conclusions rather than forcing our own priors upon it. The standard method has been to follow Fernandez et al. (2001) in assigning a 'g-prior' (Zellner, 1986) on β_k and improper priors on α the constant term and σ the error variance. Improper priors are those that are evenly distributed over their domain - complete prior uncertainty:

$$P(\alpha) \propto 1$$

$$P(\sigma) \propto \sigma^{-1}$$

$$\beta_j|\sigma, M_j \sim N(0, \sigma^2(\frac{1}{g}\boldsymbol{X_j'X_j})^{-1})$$

The key prior is the one on the regression coefficients. We assume a prior mean of 0 on the coefficients. The variance structure is defined according to Zellner's g. This means we start with an expectation of all coefficients being equal to 0 and our confidence in this prior is given by the term g. A small g represents high certainty that the coefficients are near zero whilst a large one means we are very uncertain that the coefficients are zero. A large g tends to concentrate posterior model probabilities on a few best-fitting models, known as the 'supermodel effect' (Feldkircher & Zeugner, 2009), potentially making BMA oversensitive to small changes in the data (Feldkircher & Zeugner, 2015), whilst a small g can systematically underestimate all coefficients leading to ambiguous results.

Various values for g have been suggested. A brief explanation of popular g's are given in Table 1. Unfortunately there is no consensus about which value of g is best. For a full review of different g-priors see Feldkircher and Zeugner (2009). We take the Unit Information Prior as our standard in this paper, but use the rest as a robustness test. The value of the Unit Information Prior is in its simplicity. It sets g equal to the number of observations available, thus linking our confidence in our regression results to the sample size — the information available. The Uniform Information Prior has been used widely in Bayesian model averaging such as in the papers of Feldkircher and Zeugner (2012) and Rockey and Temple (2016).

Table 1. *Descriptions of g-priors used.*

g-prior	Description
Uniform Information Prior (UIP)	$g = N$, Sets g according to the amount of information available which is the number of observations. This causes the Bayes factors to behave according to the Bayesian Information Criterion (BIC) (Kass & Wasserman, 1995).
Risk Inflation Prior (RIC)	$g = K^2$, Calibrates priors for model selection based upon the Risk Inflation (Zeugner & Feldkircher, 2015)
Benchmark Prior (BRIC)	$g = max(N, K^2)$, Fernandez et al. (2001)
Hannan-Quinn Criterion (HQ)	$g = log(N)^3$ (Zeugner & Feldkircher, 2015)
Local Empirical Bayes (EBL)	Estimates a separate g for each model using its marginal likelihood. (Zeugner & Feldkircher, 2015)
Hyper g prior (Hyper)	Uses a hyperprior distribution on g. Adjusts the posterior distribution to reflect the data's signal-to-noise ratio, reducing the sensitivity of BMA (Feldkircher & Zeugner, 2009).

In addition to providing a g-prior, we need to choose a model prior. We use binomial model priors, putting a fixed prior inclusion probability of θ on each variable which in turn determines the priors for each model. As such the prior probability of a model size k is

$$P(M_j) = \theta^{k_j}(1-\theta)^{K-k_j}$$

θ represents the inclusion probability of each regressor and k_j is the number of regressors included in M_j. In growth modeling, a model size of seven is considered standard (Barro & Sala-i-Martin, 2003; Jones & Schneider, 2005; Sala-i-Martin et al., 2004). As such we adjust θ in each set of regressions to ensure the expected model size is 7. This means $\theta = 7/K$ where K is the total number of explanatory variables in our dataset. However, when fixed regressors are used in addition to GDP per capita in the starting year, we increase the expected model size to ensure our tested variables always have the same prior probability of inclusion across tests of different fixed regressors.

Alternative model priors are possible. One approach is to make θ random as done by Ley and Steel (2009) and Bruns and Ioannidis (2020). This is to reduce the effect of possibly fixing the wrong model priors. Another approach is to set θ equal to 0.5 resulting in uniform model priors. A problem with this approach is that it centres the expected model size at $K/2$, overweighting the more numerous 'large' models which can cause overfitting. Fernandez et al. (2001) and Jones and Schneider (2006) use uniform model priors for example. Crucially, Jones and Schneider only consider models of size 7 meaning there is no overweighting of large models. Our approach of considering models of all sizes allows our results to be influenced by the strongest models regardless of their size. Given the simplicity of setting an expected model size of seven, we use this as standard practice. Uniform model priors and random model priors are used as a robustness test.

There are two other important methodological differences between our research and Jones and Schneider (2006). In their approach they use the maximum possible sample size allowed in each regression. This means the regression models are not being employed on the same observations making them less comparable. In our approach we remove observations that cannot be used in every possible regression model. Nevertheless, both approaches are affected by selection bias into the sample.

Another difference is that Jones and Schneider (2006) enumerate all their possible models and take a Bayesian average of them. They are able to do this because they keep three regressors fixed, which Sala-i-Martin (1997) considered to be particularly strong variables, ensuring there were fewer possible models.

We employ the same fixed variables only in our robustness test examining the effect of different fixed variables. We avoid this approach to ensure we do consider a wider population of possible models. To do this we only take a sample of all possible models rather than enumerating all our possible models which would take too long. To do this we use Markov Chain Monte Carlo sampler as is standard practice in the prior literature (eg. Bruns & Ioannidis, 2020; Fernández, Ley & Steel, 2001b). We use 200,000 iterations of models after a burn in phase of 100,000 iterations of models. Due to time constraints, we used a tenth of the iterations in each Bayesian model average within our bootstrapping approach which involved 1,000 Bayesian model averages. We use 20,000 iterations of models after a burn in phase of 10,000 iterations of models. By comparison, the default settings for the *BMS* package, which is designed with the application to economic growth in mind, uses 3,000 models after a burn in phase of 1,000 models.

A final note regarding methodology is necessary. In every BMA run we include the logarithm of GDP per capita in the starting year as a fixed regressor. This is because prior literature has consistently found a strong negative effect of the starting level of the logarithm of GDP per capita (eg. Barro & Sala-i-Martin, 2003, p. 496, p. 521) dubbed the advantage of backwardness. This has theoretical roots in neoclassical growth models in which diminishing marginal returns to investment allows poorer countries to grow faster. Given the high correlation of IQ and other variables with GDP in the starting year, any models that do not include GDP would bias national IQ's coefficient downwards as it takes on some of the effect of the advantage of backwardness. Given this potentially large co-dependency of GDP per capita and other explanatory variables, we follow common practice by including it as a fixed regressor.

Section 4: Data

National IQ

National intelligence can be measured in two ways. One way is the psychometric method, administering IQ tests to more-or-less representative samples within each country. These results were collected by Richard Lynn and co-authors (Lynn & Vanhanen, 2002, 2012; Lynn & Becker, 2019). These scores are adjusted for changes in IQ over time, the Flynn effect, assuming these changes are linear and the same across countries. The UK's score is then set to 100, and one standard deviation in IQ amongst British people is set to 15. This is the 'Greenwich mean IQ'.

The Lynn national IQ data derived from IQ tests have been criticized by Wicherts et al. (2010a,b). The critics suggest that scores from Sub-Saharan countries are implausibly low, do not use representative samples, and may be unduly deflated by cultural factors, poor nutrition, and poor education. Lynn and Meisenberg (2010a) have responded to these critiques defending the quality of the national IQ scores. Lynn and Meisenberg argued many of the studies Wicherts et al. used to show higher IQ for Sub-Saharan Africans are unrepresentative because they are based on university students. David Becker later went back to the sources and recalculated the national IQs with a more rigorous set of conditions for inclusion (Lynn & Becker, 2019). These scores correlated very well with Lynn and Vanhanen's (2002) original IQ scores and support many of the very low scores for the most underdeveloped nations. Moreover, Lynn and Vanhanen's (2002) IQ scores for Sub-Saharan Africans are consistent with the results of the Southern and Eastern Africa Consortium for Measuring Education Quality (SACMEQ) assessment (Sandefur, 2018; Thompson, 2016).

Sample representativeness is a serious concern for using national IQs as a measure of human capital. Other critiques are not necessarily so important for economics research. Whilst psychologists may be concerned that culture, nutrition and education quality may mask a nation's 'true' or potential intelligence, economists are interested in the actual, phenotypic ability that determines the cognitive human capital of workers. These environmental factors may be important for reverse causality, which we discuss in the Causality section of the paper.

To avoid specific critiques based on the psychometric national IQ datasets, we also repeat our analysis using a second set of cognitive measures: student assessment tests. This includes the PISA, TIMSS and PIRLS tests which are regularly given to students in a large range of countries testing proficiency in mathematics, the native language, and science. As a measure of educational output without pretensions of measuring 'IQ', these scores avoid criticism that they do not in fact measure intelligence. Nonetheless, they appear to measure the same cognitive human capital that IQ tests measure because their scores correlate highly ($r > .9$) with different updates of Lynn's national IQs (Lynn & Meisenberg, 2010b; Meisenberg & Lynn, 2011). This strong relationship between test scores collected by intergovernmental organizations and Lynn's psychometric national IQs strongly supports the validity of psychometric IQs. Furthermore, student assessments have been popular amongst economists, having been employed in various research (Hanushek & Woessmann, 2012, 2015) and used to create human capital measures for the World Bank (Angrist et al., 2019, 2021; Lim et al., 2018).

Student assessment tests were available at the time Jones and Schneider (2006) performed their Bayesian model averaging, but sample sizes were very small. Since then, many more countries have had their school children take part in these standardized assessments, allowing larger samples to be used in the study of economic growth (e.g. Hanushek & Woessmann, 2015). Thus in our replication of Jones and Schneider, we employ various measures of cognitive human capital derived from student assessments to avoid criticism unique to Lynn's national IQ data. Furthermore, given the discussed sensitivity of Bayesian model averaging, it is important to see whether minor variations in national IQ data could alter the results.

Table 2. *National IQ data.*

Name	Citation	Notes
Becker, psychometric	Lynn & Becker (2019)	Column 'E' of the 'FAV' tab, version 1.3.3 of the national IQ dataset (https://viewoniq.org/). This variable recreates Lynn & Vanhanen's psychometric IQs, with different methodology and selection criteria, weighting samples by their quality and size.
Becker, psychometric/SAS	Lynn & Becker (2019).	A simple average of Becker's psychometric and student assessment IQ scores.
Becker, SAS	Lynn & Becker (2019)	Mean IQs calculated from PISA, TIMSS and PIRLS tests.
Hanushek & Woessmann, SAS	Hanushek & Woessmann (2012).	A 'cognitive skills' measure created from 12 different student assessment tests from 1964-2003.
L&V 2002, Psychometric	Lynn & Vanhanen (2002)	The original set of psychometric national IQs, standardized to British IQ, adjusting for a constant Flynn effect
L&V 2012, Psychometric	Lynn & Vanhanen (2012)	Updates Lynn & Vanhanen's (2002) national IQ scores with additional samples and countries.
Rindermann, Psychometric/SAS	Rindermann (2018)	Weighted average of psychometric scores and student assessment scores, putting twice as much weight on SAS, missing data extrapolated from International Mathematics Olympiad
Rindermann, SAS	Rindermann (2018)	Student assessment scores from a range of tests standardized across tests and time.
World Bank HLOs (Harmonized Learning Outcomes), SAS	Angrist *et al.* (2021)	Student assessment results standardized across different tests. World Bank does not provide an average of these scores across time so we only used HLOs from 2015. https://datacatalog.worldbank.org/dataset/harmonized-learning-outcomes-hlo-database

For this study, we picked a wide range of publicly available popular psychometric, student assessment (SAS) and mixed measures of national IQ. A list and brief description of these variables are presented in Table 2. Our standard national IQ is the mixed SAS and psychometric score of Rindermann (2018). This measure has the largest sample size of all the national IQs because it uses a wide range of student assessments and national IQs. This measure also ensures a larger sample size for each country making IQ estimates more accurate. Finally, the data puts a greater weight (3 to 1) on student assessment scores due to their larger sample sizes, making it preferable to mixed national IQ scores from Becker's national IQ dataset. The rest of our selected national IQs are used as a robustness test.

A correlation matrix for our national IQ variables is shown in Table 3. As in the prior literature, cognitive skills whether measured via IQ tests, student assessments or a mixture correlate extremely well. The lowest correlation seen between any two measures is .83.

Table 3. *National IQ correlation matrix; P = psychometric, S = school achievement, H&W = Hanushek & Woessmann, L&V = Lynn & Vanhanen, R = Rindermann, HLO = Harmonized Learning Outcomes (World Bank), mean* r = *mean correlation with the other measures.*

	Becker P&S	Becker S	H&W, S	L&V 2002 P	L&V 2012 P	R, P&S	R, S	HLO, S	mean *r*
Becker, P	0.95	0.86	0.83	0.83	0.88	0.89	0.87	0.84	0.87
Becker, P&S	1.00	0.98	0.92	0.92	0.96	0.97	0.96	0.93	0.95
Becker, S		1.00	0.93	0.92	0.96	0.97	0.97	0.94	0.94
H&W, S			1.00	0.87	0.92	0.95	0.95	0.89	0.91
L&V 2002, P				1.00	0.95	0.94	0.92	0.88	0.90
L&V 2012, P					1.00	0.98	0.96	0.94	0.94
R, P&S						1.00	0.99	0.94	0.95
R, S							1.00	0.92	0.94
HLO, S								1.00	0.91

Control variables

We employ two different datasets of control variables in this paper. The first is the dataset created by Sala-i-Martin, Doppelhofer and Miller (2004) used in their Bayesian modeling, hereinafter named SDM. SDM first found variables that predicted economic growth in the prior literature. From this group, only select variables were included in the dataset if they had values sufficiently close to the year 1960. Unfortunately, some variables start in 1965. This was to reduce the problem of endogeneity. To ensure a large sample size, SDM then selected variables that would maximize the product of the sample size and the number of variables. This left SDM with 67 explanatory variables.

SDM studied the determinants of growth for the period 1960-1996. To incorporate more recent GDP data, we employ the growth rate from 1960-2010 and use new figures for log GDP per capita. SDM's GDP data came from version 6 of the Penn World Tables (PWT) whereas ours come from PWT 10 (Feenstra et al., 2015), variable RGDPe, to provide log GDP per capita in the starting year and the growth period. RGDPe is expenditure-side real GDP at chained PPPs, to compare relative living standards across countries and over time.

The SDM dataset has two key advantages. The first is its large number of control variables and large sample size. The second advantage is that it is a well-tested and well-respected dataset having been used in many subsequent papers using Bayesian modeling to study economic growth (eg. Ciccone et al., 2010; Doppelhofer & Weeks, 2009). This limits criticism regarding our choice of control variables and ensures we could not commit 'data dredging' — specifically designing a dataset to prove our hypotheses.

We could have used the dataset created by Sala-i-Martin (1997a,b) or the subsection of it which was first used for Bayesian modeling by Fernandez et al. (2001). The subsection takes only variables found to be 'important' and well-performing in Sala-i-Martin's (1997a,b) model averaging method and other variables from the Sala-i-Martin data which did not reduce the sample size. This subsection is also a popular dataset used in other papers on Bayesian modelling (e.g. Horvarth, 2011; Ley & Steel, 2007). Jones and Schneider (2006) use a subset of the Sala-i-Martin dataset, in a similar fashion to Fernandez et al. (2001), but take only the significant variables.

We think the SDM dataset is superior to the Fernandez et al. and Jones and Schneider dataset. Excluding plausible control variables, which model averaging has previously found little support for, would undermine our goal of testing national IQ against all plausible variables and theories. The rejection of the excluded variables could represent a false negative, which may perform well under different specifications.

The second dataset we use is the dataset used by Bruns and Ioannidis (2020), which is itself intended to recreate a dataset similar to SDM. Hereafter the modified Bruns and Ioannidis data is labelled the BI dataset. In their paper, Bruns and Ioannidis test the stability of predictors across different time frames. This restricts their variables to those that are available regularly for different periods. This property of the dataset makes it easy for us to test the effect of national IQ in different time periods from 1960 to 2010: 20-year time periods starting in years divisible by 5. When we use this dataset, all control variables except national IQ are taken from the starting year of the growth period. This is a substantial limitation of our method which is discussed further in the Causality section.

Furthermore, Bruns and Ioannidis take many of their variables from recent editions of the Penn World Tables. This allows us to easily perform a sensitivity test using different editions of the Penn World Tables. Bruns and Ioannidis attempt to further remove any endogeneity by removing variables describing countries in the duration of the growth period, such as a socialist dummy, average inflation rate, and proportion of time the country spends at war. Despite this, Bruns and Ioannidis still employ a variable measuring growth in the terms of trade (that is change in exports divided by imports), which we remove in our dataset.

Ioannidis and Bruns use time-invariant geographic variables used by SDM, but which originally come from the Gallup et al. (2001) geography dataset. Because the geographic variables contain the same missing variables, it would not reduce our sample size to include additional ones. As such, to test national IQ against a larger body of variables and thus plausible theories, we include additional time-invariant geographic variables from the Gallup et al. (2001) geography dataset.

A full list of all the variables employed in these datasets is given in Tables 11 and 12 of the Appendix.

Additional explanatory variables

To impose discipline on the estimation methodology, we have not added any additional control variables to the SDM dataset, except NIQ. But we also do this because these datasets have been optimized to maximize the sample size. If we carelessly add additional variables, the efficiency of our Bayesian model averaging could be severely affected. With Rindermann's national IQ, the SDM dataset has 82 observations and the BI dataset has 43 observations. These are not sample sizes that should be decreased further unnecessarily. Thus to test the effect of additional variables we use these as a single and separate test. Additional variables included are given in Table 4.

The first type of variables we add are psychological ones. These are social trust, time preference, and kinship intensity. Social trust is measured by how a nation's population responds to the following question from the World Values Survey: "Generally speaking would you say that most people can be trusted or that you can't be too careful in dealing with people?" Many economists have found a positive relationship between a nation's level of social trust and their GDP, see Bjørnskov (2017) for a review of this literature. A trusting, or perhaps a trustworthy, society can reduce the transaction costs of business and reduces the risk of theft and rent-seeking in business (Algan & Cahuc, 2010, 2013). Intelligent individuals (e.g. Carl & Billari, 2014) and intelligent societies (Rindermann, 2008) tend to be more trusting. This suggests social trust might mediate national IQ's effect or that it might confound IQ's effect on economic growth. Roth (2009) found that with national fixed effects, social trust had a negative association with economic growth. Later, Carl (2014) found social trust no longer predicted GDP when IQ was controlled for. We replicate this test by including social trust as a variable.

Table 4. *Extra variables.*

Variable name	Source and description
Social trust	Measure of self-reported social trust; from Carl (2014) and derived from the World Values Survey.
Time preference	Time preference measure derived from surveys and correlated variables such as credit risk; from Rieger et al. (2021).
Kinship Intensity Index	Measure of 'kinship intensity' from Schulz et al. (2019)
Economic freedom	Fraser Institute's Economic Freedom of the World index (Murphy & Lawson, 2018)
UV radiation	Taken from Andersen et al. (2021)

Time preference or patience refers to how individuals value consumption across different time periods. Less intelligent individuals tend to have a larger time preference (Mischel et al., 1972, 1989; Watts et al., 2018), preferring smaller rewards today over larger ones in the future. At the level of nations, time preference has been estimated using the Global Preference Survey (Falk et al., 2018) in which individuals are asked about how they would make trade-offs between cash prizes given at different points in time. Time preference may influence economic growth through higher savings and investment (Jones, 2010, 2012). Unsurprisingly, national IQ correlates with savings rates (Jones, 2010) and time preference measures (Kirkegaard & Karlin, 2020) at the national level. When Karlin and Kirkegaard tested national IQ and time preference as predictors of national welfare, they found time preference was statistically insignificant when IQ was included. Like social trust, time preference represents a potential mediator or confound of national IQ so we include it as an additional variable.

Joseph Henrich (2020) has suggested that marriage patterns have played a key role in determining the prosperity of the West. Rules created by the Catholic Church discouraged Europeans from marrying their relatives, and therefore exposure to the Catholic Church is associated with non-cognitive psychological differences today (Schulz et al., 2019). This reduction of inbreeding is thought to have reduced the intensity of kin-based institutions and allowed for individualism which has driven innovation and capitalism. Henrich (2020) found individualism to be associated with patents per capita. We employ the Kinship Intensity Index, which measures the presence of cousin-marriage preferences, polygamy, co-residence of extended families, clan organization, and community endogamy (Schulz et al., 2019). As a novel variable to indicate psychological differences, it is a natural rival to national IQ.

As mentioned in Section 2, smarter individuals tend to support a free market (Carl, 2014a, 2015; Kirkegaard et al., 2017). Smarter nations tend to have a freer economy which may mediate national IQ's effect on GDP (Christainsen, 2020; Rindermann & Thompson, 2011). The SDM dataset includes an old "Degree of Capitalism" measure (Hall & Jones, 1999). However, the variable is not very sophisticated (Christainsen, 2020) and is ordinal rather than continuous. To improve upon this we employ the Economic Freedom Index created by the Fraser Institute (Murphy & Lawson, 2018). In our test of additional variables, we add this to the BI dataset and replace the Degree of Capitalism variable with the Economic Freedom Index in the SDM dataset.

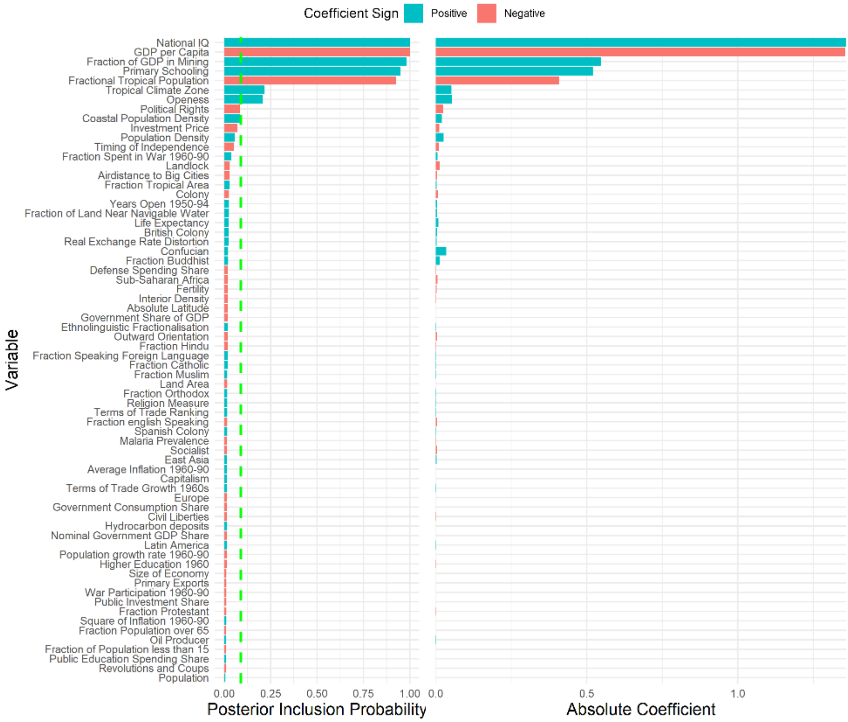
National variations in cognitive ability have often been explained as an evolutionary adaptation to the challenges of cold winters (Frost, 2019; Lynn, 1987; Rushton, 1995). The theory fits the data with groups further from the equator having larger cranial capacity (Kanazawa, 2008) and skin color having the strongest relationship, of all variables, with national IQ (Templer & Arikawa, 2006). We discuss Cold Winters theory in greater depth in the Causality section. If the theory is wrong, national IQ might only predict growth due to geographic confounding. After all, many economists have found strong relationships between absolute latitude (distance from the equator) and GDP per capita (e.g. Nordhaus, 2006). Absolute latitude is already in the SDM and BI datasets, but to better test IQ against possible climatic confounding, we also include UV radiation as one of our extra variables.

Section 5: Results

Before using any of our robustness tests or additional variables, we ran Bayesian model averaging with the BI and SDM datasets with Rindermann's national IQ scores. As expected, national IQ performs extremely well with a posterior inclusion probability of 1. Its absolute coefficient is the largest of all variables tested, not including log GDP per capita in 1960 because it is used as a fixed regressor in all regressions. The SDM dataset has the larger number of explanatory variables (81) and observations (69), whereas our BI dataset has 63 variables and 63 observations.

In the SDM dataset (Figure 2), national IQ has an absolute coefficient of 1.4, implying a one standard deviation increase in national IQ is associated with a 1.4 percentage-point increase in the growth rate. In the BI dataset, national IQ has a somewhat smaller coefficient of 1.2. Without standardizing our independent variables, one national IQ point is associated with 0.09 percentage point higher growth per year with the BI data and 0.11 percentage points in the SDM dataset. If we interpret our results through the framework of an exogenous model, each additional national IQ point is associated with a 6.5% larger GDP per capita in the BI data and 7.8% in the SDM data. This compares with a previous estimate of 6.4% from Jones and Schneider (2006). Details of these calculations and discussion of different growth models (exogenous, endogenous and the Nelson-Phelps technology diffusion model) can be found in the online supplement.

Some other variables do perform well, with posterior inclusion probabilities higher than their prior probabilities. In the SDM dataset there are five other variables that pass this test: fraction of GDP in mining, primary school enrolment, fraction of population living in the tropics, fraction of the country in the tropics, trade openness. However, only the first three of these variables had posterior inclusion probability greater than 0.5 indicating they are more likely to be included in the best model than not. Moreover even of these variables the highest coefficient is less than half of national IQ's coefficient indicating IQ is substantially more important than even best performing competitors.

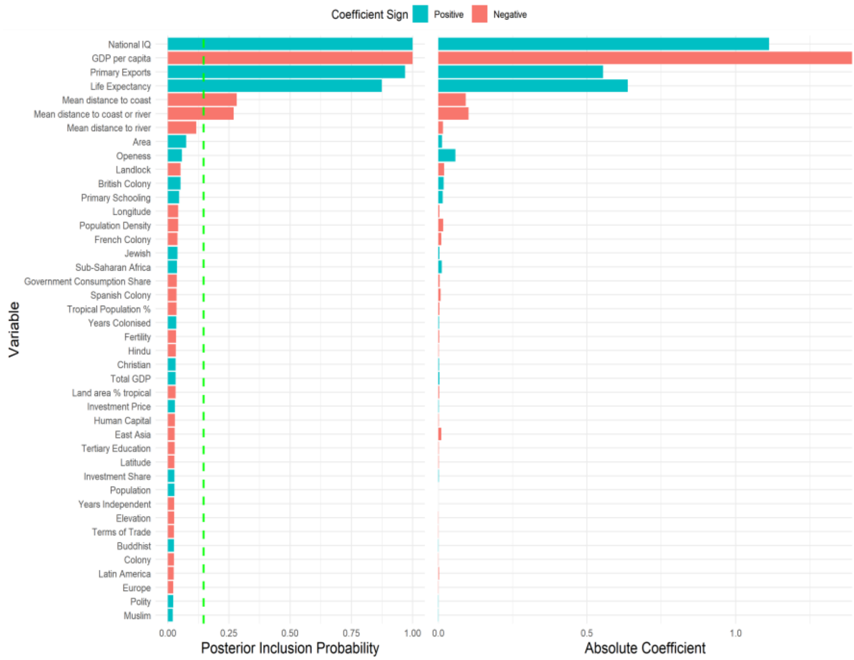


Note: Dashed line denotes the prior inclusion probability

Figure 2. *SDM data main results*

The BI dataset has four other variables with higher PIPs than priors (Figure 3). These are life expectancy, exports of primary goods as a percentage of GDP, average distance to rivers, and average distance to the coast. The failure of primary school enrolment to have a higher PIP than prior in the BI dataset should make us skeptical of its robustness despite its good performance in the SDM dataset.

An important difference between our results and Sala-i-Martin et al.'s (2004) results is that their East Asian dummy had the highest PIP, and fraction Confucian was the 9th best variable. These variables do not perform well in our results. This is likely due to the fact East Asian countries have high national IQs, and Sala-i-Martin et al.'s variables were mostly proxies for high IQ. Likewise the Sub-Saharan African dummy variable was the 10th best variable for Sala-i-Martin et al. (2004). This suggests much of the regional effects found in prior studies might be spurious due to being confounded with national IQ.



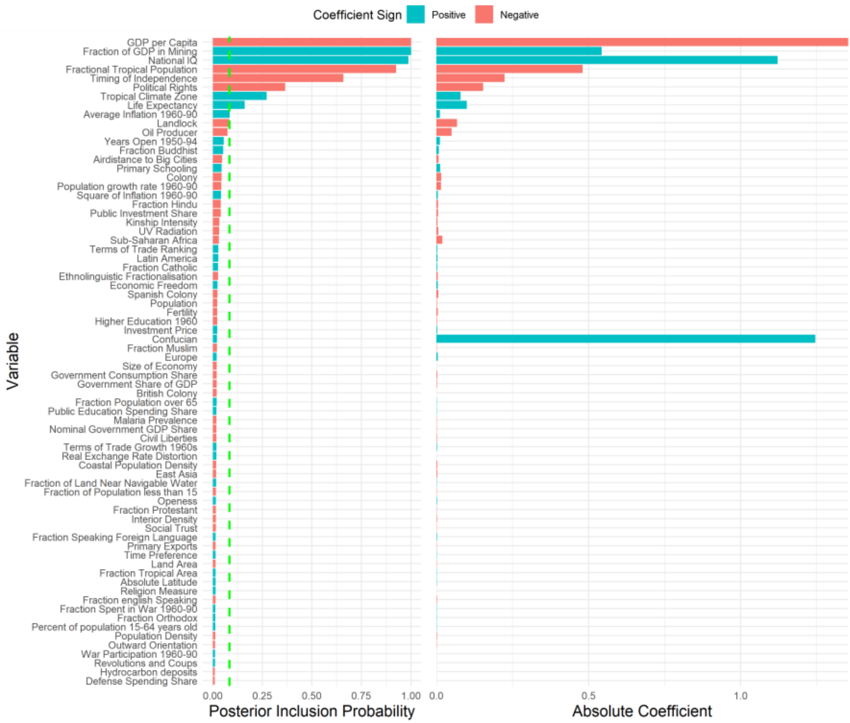
Note: Dashed line denotes the prior inclusion probability

Figure 3. *BI data main results.*

Results with additional variables

To test whether national IQ's apparent success is due to omitted variable bias, we ran the same Bayesian model averaging but with popular variables that are related to or confounded with intelligence as represented in Table 4. National IQ's posterior inclusion probability falls from 1 to 0.96 in both datasets. The results have barely changed indicating that national IQ's predictive powers cannot be explained by possible confounds.

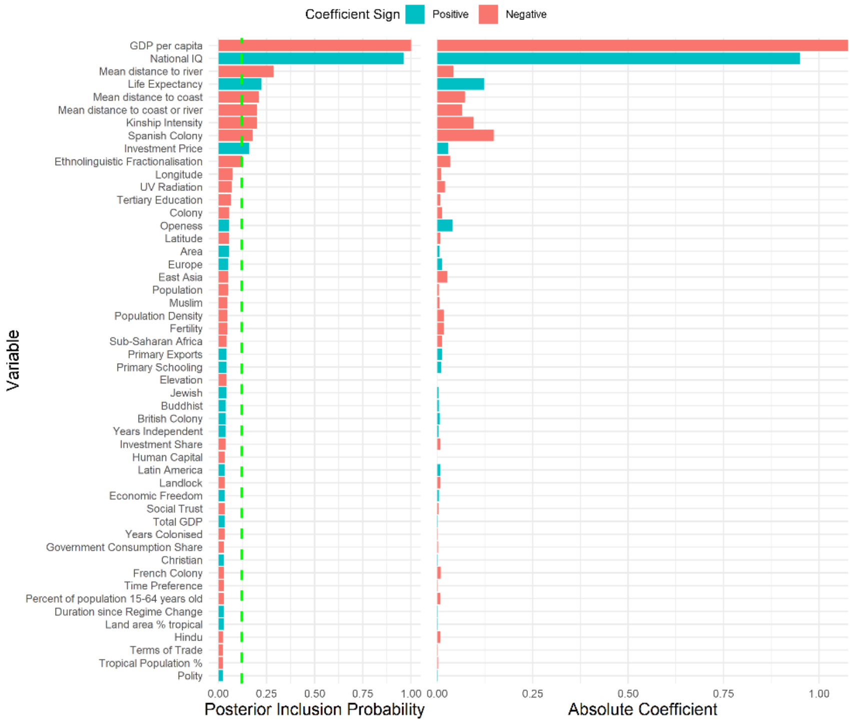
UV radiation, time preference, social trust, and economic freedom did not have a higher PIP than prior (Figures 4 & 5). This replicates the findings of Carl (2014) and Kirkegaard and Karlin (2020) who respectively tested whether social trust and time preference could explain national IQ's relationship with economic growth. The failure of UV radiation and latitude to robustly predict economic growth suggests that IQ's relationship with growth is not due to spatial autocorrelation. The Fraser Institute's Economic Freedom Index's PIP is lower than its prior in both datasets. This contrasts with Christainsen (2020) and Weede and Kämpf (2002), who find that economic freedom is statistically significant in their growth regressions which use national IQ. Nonetheless, this result is not unexpected because degree of capitalism has shown inconsistent results in studies using rigorous model averaging methods. For example, Sala-i-Martin (1997) found that the variable was robust whilst Sala-i-Martin et al. (2004) found the degree of capitalism to have the third lowest PIP out of 67 tested variables.



Note: Dashed line denotes the prior inclusion probability

Figure 4. *SDM results with extra variables.*

Only one of our extra variables, the Kinship Intensity Index, has a higher posterior than prior probability. Moreover this result only occurs in the BI dataset. It has a posterior inclusion probability of 0.20 which is not large but double the prior inclusion probability. Given that we have added four extra variables into two different datasets, we should probably expect at least one of the variables to perform well once by chance. Nonetheless, the result is consistent with Joseph Henrich's hypothesis that societies with lower kinship intensity tend to become more prosperous (Henrich, 2020).



Note: Dashed line denotes the prior inclusion probability

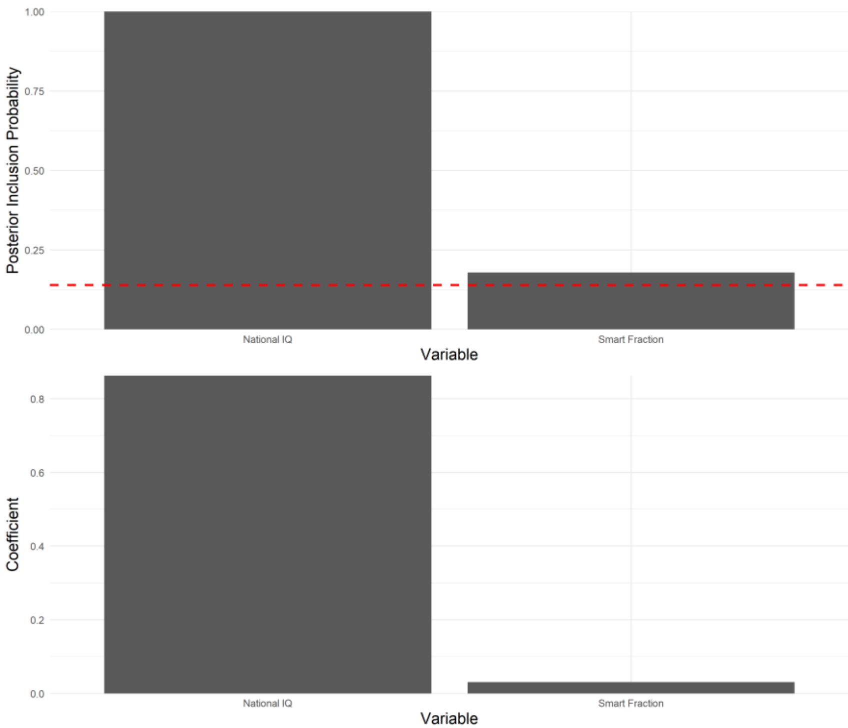
Figure 5. *BI results with extra variables*

Smart fraction theory

So far we have only considered the effect of average national IQ on economic growth, however national populations do not just differ in the means but in their entire distributions. An important question is whether the smartest fraction of a country's IQ distribution plays a more important role than the mean.

This is a plausible theory on many grounds. Psychologists such as Terman (1916) and Jensen (1980, p. 114) have argued that because exceptional achievement is created by the brightest, they will have the largest effect on societies. Many previous studies looking at economic development and economic growth have found larger effects of the 95th percentile IQ than the 5th percentile or the mean level (Rindermann, 2012, 2018; Rindermann, Kodila-Tedika & Christainsen, 2015).

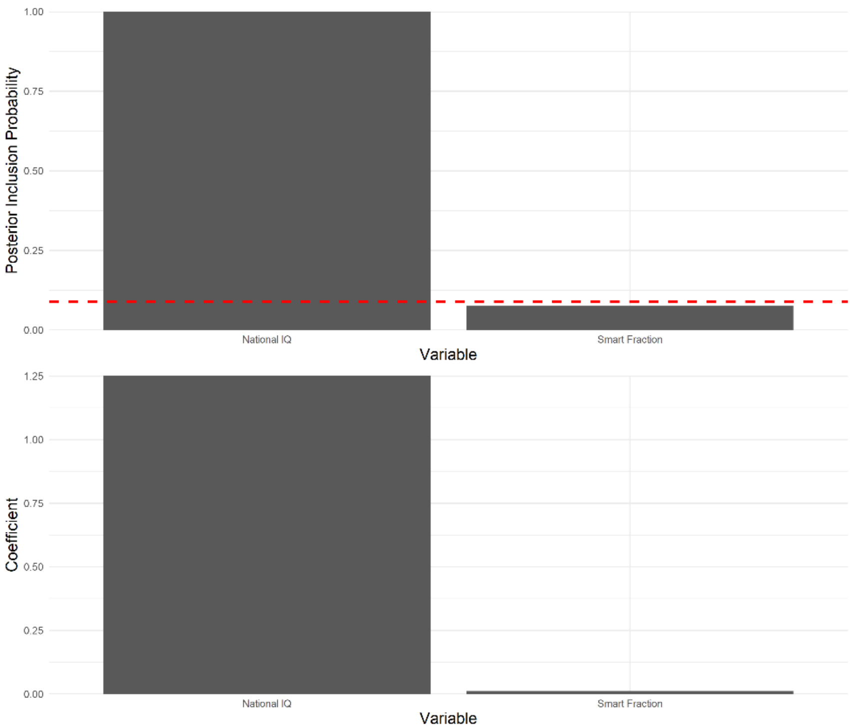
A particular problem with this research is that the mean IQ and IQ of the top 5% correlate highly, even at r = .98 in Rindermann, Kodila-Tedika and Christainsen (2015). With such multicollinearity it is difficult to differentiate the effect of the average IQ and that of the cognitive elite. Kirkegaard (forthcoming) responds to this issue by regressing the top 5% IQ on the mean IQ and taking the residuals as a measure of how smart the brightest in the nation are compared to what one would expect based on the mean ability. In regressions both the mean and residualized elite IQs were statistically significant predictors of national welfare across many variables.



Note: Dashed line denotes the prior inclusion probability

Figure 6. *Smart fraction results using BI data.*

We test Kirkegaard's smart fraction measure in the SDM and BI datasets (Figures 6 & 7). In this test we use Rindermann's (2018) student assessment IQ scores rather than his combined psychometric and student assessment scores because the measurement of the nation's smartest 5% of students comes only from the results of student assessments.



Note: Dashed line denotes the prior inclusion probability

Figure 7. *Smart fraction results using SDM data.*

In general the Smart Fraction variable performed poorly, with a PIP of 0.18 in the BI dataset and 0.08 in the SDM dataset. The result for the BI dataset was marginally above the prior inclusion probability of 0.14. Moreover, the estimated coefficient in the BI dataset was 0.03, implying a standard deviation increase in the elite IQ, over and above what is expected from the mean IQ, only increases economic growth by 0.03% a year. Despite performing well in a few regression models of the general socioeconomic factor for nations (Kirkegaard and Carl, in review), we find smart fraction theory has little explanatory power for economic growth. Kirkegaard and Carl (in review) note that measurement error is largest in the tails of a distribution, suggesting our measure of the smart fraction may be weak. Use of residuals to test smart fraction theory should be re-evaluated in the future when further samples allow better estimations.

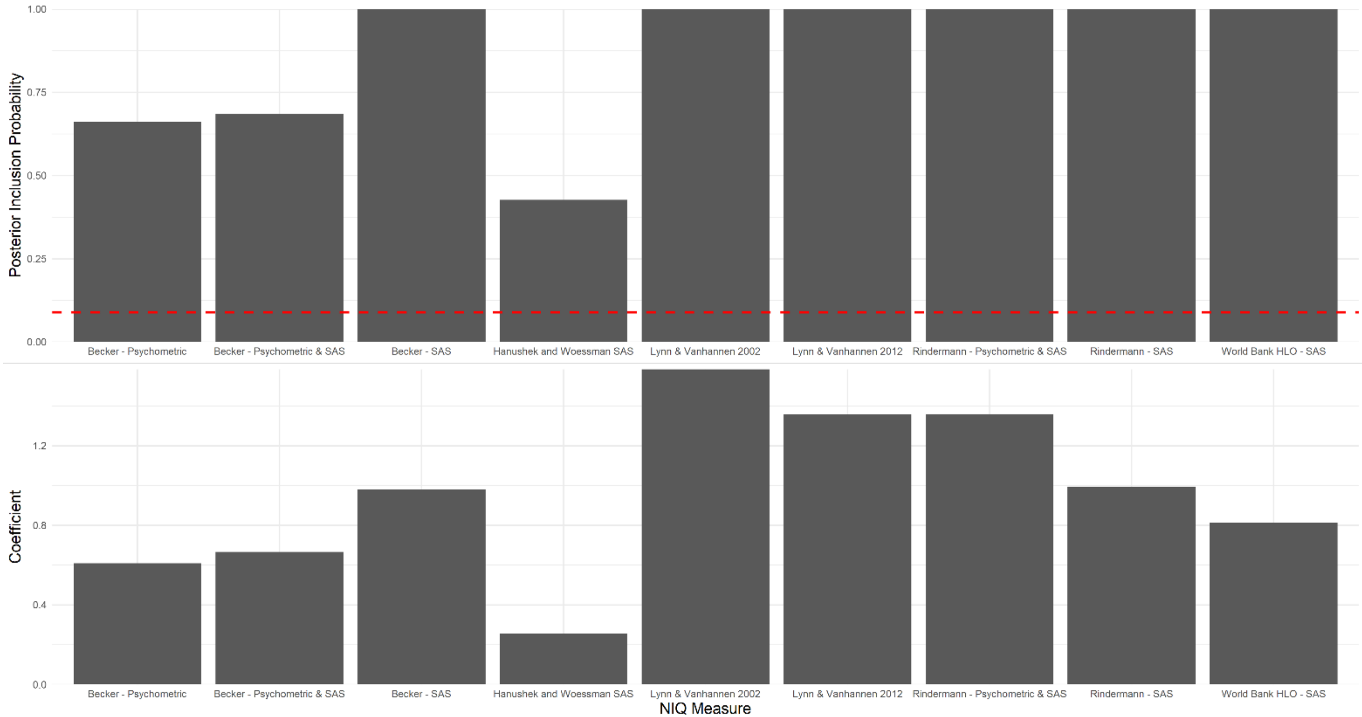
A more general problem for smart fraction theory is that it is not clear how the smart fraction should be defined. In the work of Rindermann (2018) the smart fraction is defined as how smart the smartest 5% of a country is. However, in one formulation of smart fraction theory (La Griffe du Leon, 2002) the variable of concern is the fraction of the population of a 'smart' IQ of 105 necessary for complex work. Our residualized approach compounds this problem. If the IQ of the top 5% is higher than what would be expected of the mean IQ, then our measure also tells us that the median IQ is lower than the mean. In other words, our residualized approach to measuring Rindermann's conception of the smart fraction may be negatively related to La Griffe du Lion's measure of the smart fraction. An alternative operationalization of smart fraction theory may be necessary.

Alternate national IQs

So far we have used Rindermann's (2018) national IQs as our standard. This was because by combining student assessment scores with national IQs we attain data on more countries and have larger samples for the countries included. Here we try different national IQs. Given the sensitivity of Bayesian model averaging it is possible that even slight changes in national IQ scores due to measurement issues could create substantial differences in results.

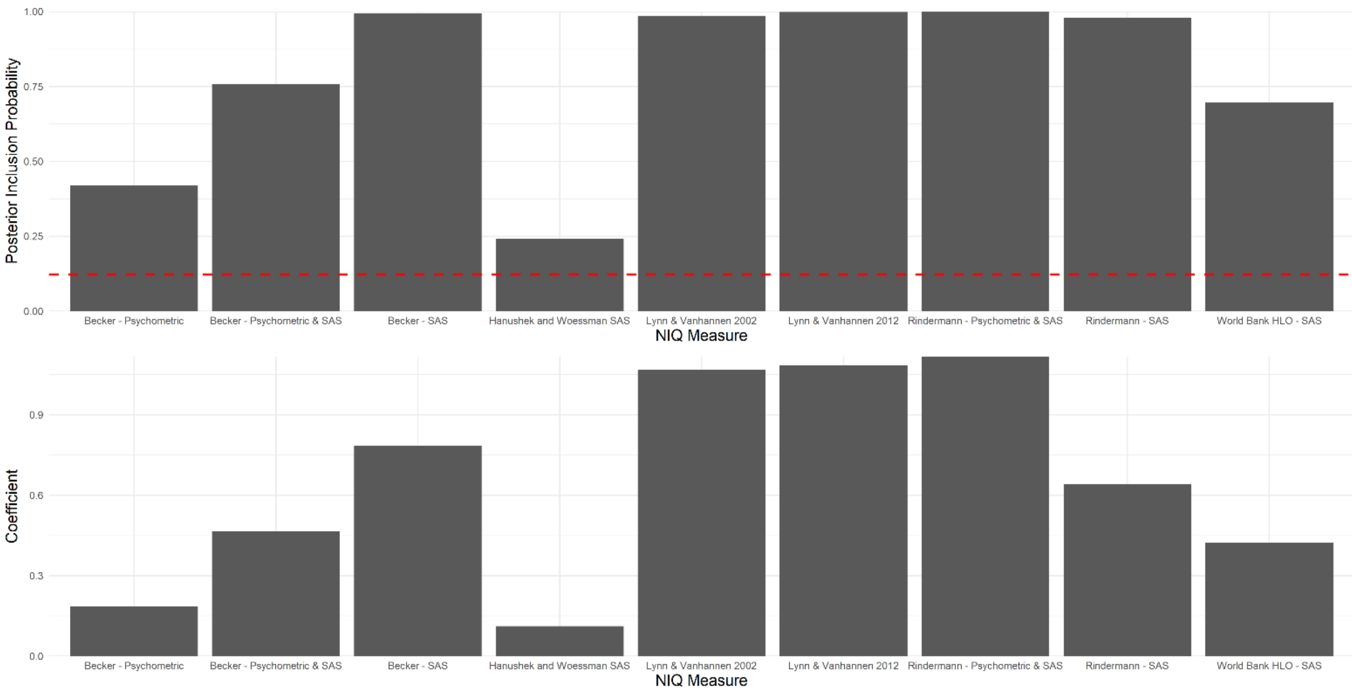
In the SDM dataset, the posterior inclusion probability is greater than the prior probability for every national IQ score employed, with two thirds performing extremely well with a PIP equal to one (Figure 8).

There is no obvious difference in the PIPs between student assessment scores such as Hanushek and Woessmann's (2012) and the World Bank's Harmonised Learning Outcomes, and the psychometric IQ scores such as Lynn and Vanhannen's (2002, 2012). Nonetheless the coefficients for the student assessment scores tend to be lower, although the situation is reversed in David Becker's data with student assessments outperforming psychometric IQ. It should be noted that the observations available were different for each IQ score meaning the coefficients are not perfectly comparable. In the BI dataset, national IQ's coefficients and PIPs are smaller, but still robust (Figure 9). This should be expected given that even with the use of Rindermann's psychometric and SAS national IQs, the sample size is only 43.



Note: Dashed line denotes the prior inclusion probability

Figure 8. *Results of different NIQs using SDM data.*



Note: Dashed line denotes the prior inclusion probability

Figure 9. *Results of different NIQs using BI data*

.

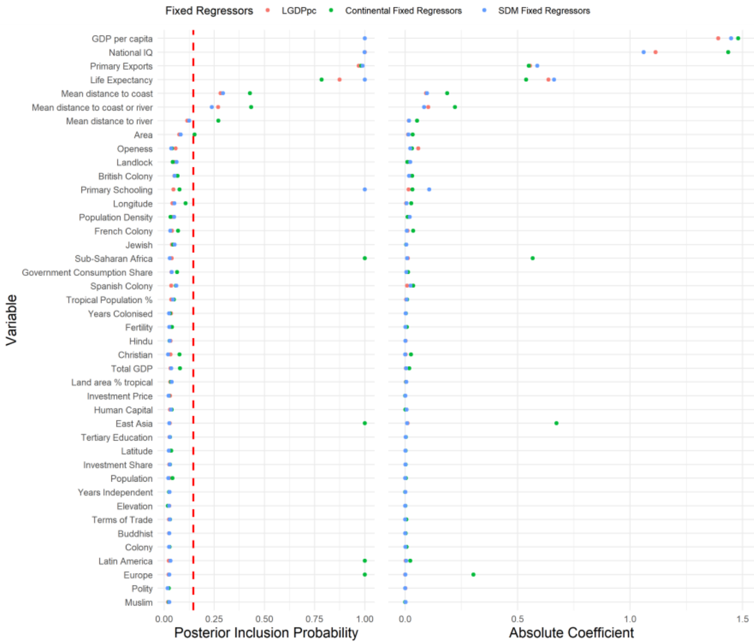
The IQ data with the largest coefficient on growth is also the oldest — Lynn and Vanhanen's (2002) dataset. This is perhaps particularly surprising given the scale of criticism this data has been given methodologically. In particular critics suggested the scores could be biased against Sub-Saharan African countries because their scores were so low. However, in our results Sub-Saharan Africa has a negative coefficient on growth. The complete results are available in the online supplement. If Sub-Saharan Africa was not inherently bad for growth and Richard Lynn's national IQ scores underestimated the human capital of Sub-Saharan Africans, then the Sub-Saharan African binary variable should have a positive coefficient. However, the estimated coefficient is negative. These results support Garett Jones's (2012) comments on Lynn's national IQ scores "If national average IQ estimates are indeed 'biased', they appear to be biased in favor of productivity growth."

Penn World Tables

Ciccone and Jarociński (2010) found that using different editions of the Penn World Tables leads to radically different results, suggesting the relevant economic data is simply not good enough to reliably find the best explanatory variables. Using national IQ as an explanatory variable with the BI dataset we replicate their test. The BI dataset is used because a substantial number of its variables come from recent editions of the Penn World Table allowing us to easily use different versions of the same variables.

Across the five most recent editions of the Penn World Tables (PWT) we find remarkably similar results. National IQ has a PIP of 1 regardless of what Penn World Table is used. National IQ's coefficient ranges from 1.1 to 1.3 (Figures 10 & 11). The coefficient is typically larger in older versions of the Penn World Tables. This is consistent with prior research finding that older editions of the Penn World Tables tend to provide more accurate measures of GDP, with stronger correlations to proxies such as light intensity (Johnson et al., 2013; Pinkovskiy & Sala-i-Martin, 2016). This may imply that better measures of GDP would further increase our estimate of national IQ's effect.

The consistency of national IQ in the face of measurement error and different observations from different editions of the Penn World Tables further supports the idea that national IQ is an extremely robust predictor of economic growth.



Note: Dashed line denotes the prior inclusion probability

Figure 10. *Results of different PWTs using BI data.*

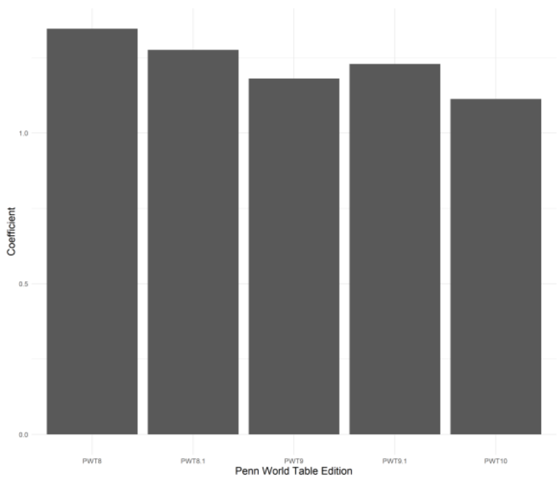


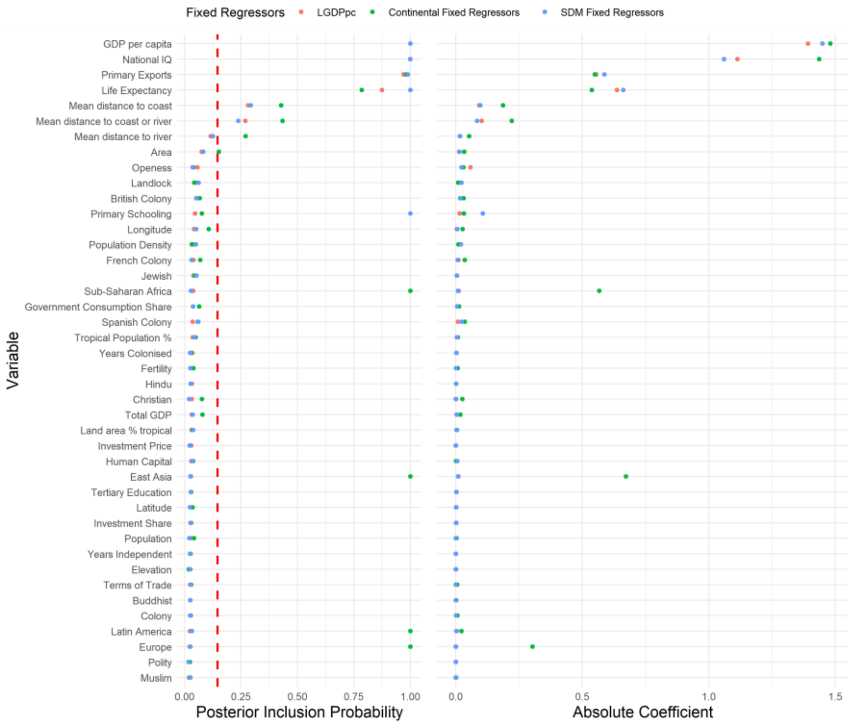
Figure 11. *National IQ's coefficient on economic growth using different PWT data.*

Fixed regressors

Employing regional dummies as fixed regressors has been found to reduce the sensitivity of Bayesian model averaging to measurement error in the Penn World Tables (Rockey & Temple, 2016). Many explanatory variables show greater variation across regions than within them. This means that regressions with or without different regional dummies can find radically different results for many explanatory variables. Because of this, Rockey and Temple (2016) recommend using regional dummy variables alongside the logarithm of GDP per capita as fixed regressors. This may be particularly important to diminish omitted variable bias for national IQ given there are large regional variations in IQ scores.

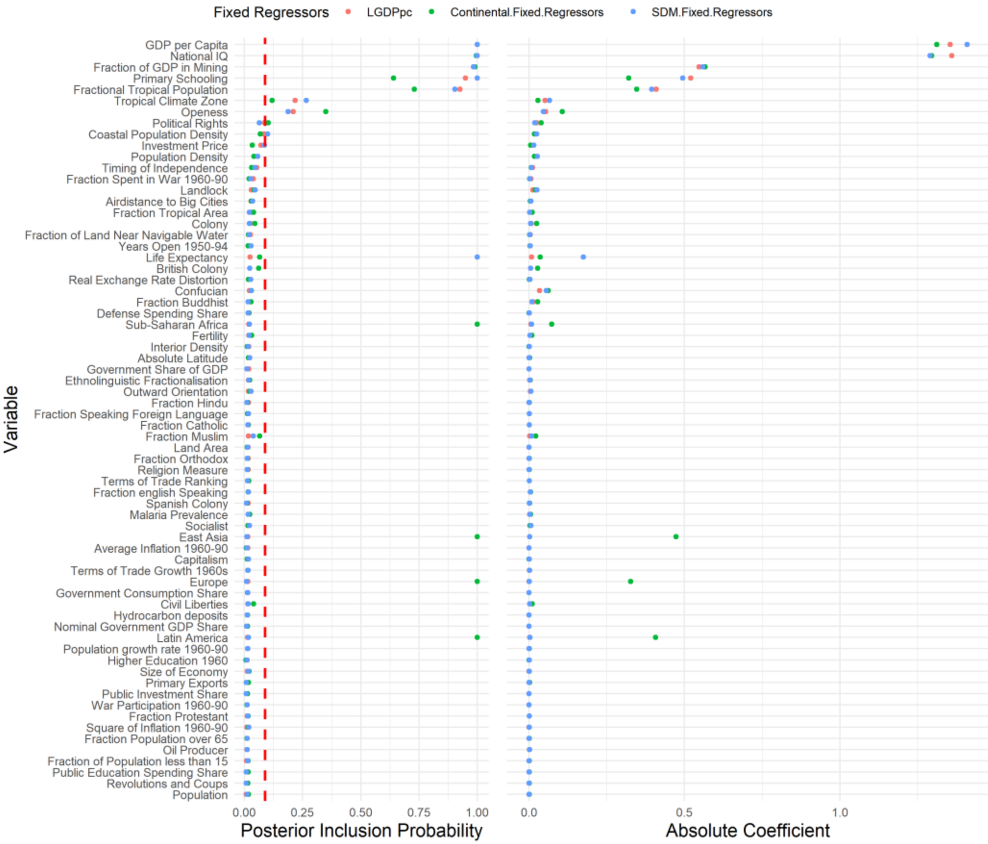
We reran our Bayesian model averaging with regional dummies in every regression. We also tried the fixed regressors utilized by Sala-i-Martin (1997) and Jones and Schneider (2006). These are primary school enrolment in 1960, life expectancy in 1960, and the logarithm of GDP per capita in 1960. We do this not only to allow our results to be comparable with Jones and Schneider's, but also to stress test national IQ in case it is confounded with life expectancy or education, showing co-dependency with these variables. After all, better health and education may increase national intelligence.

In all variations of fixed regressors, national IQ still has a posterior inclusion probability of 1 (Figures 12 & 13). Although national IQ's coefficient is larger when regional dummies are used in the BI dataset, it is slightly smaller under this scenario in the SDM dataset. For the use of Sala-i-Martin's fixed regressors, the situation is reversed with a larger coefficient in the SDM dataset and a lower one in the BI dataset. We can conclude that fixed regressors do not substantially alter our results. Moreover, the effect of IQ is not due to regional confounding.



Note: Dashed line denotes the prior inclusion probability. LGDPpc fixed regressors include log GDP per capita. Continental fixed regressors include log GDP per capita and regional dummies. SDM fixed regressors include life expectancy, primary school enrolment and log GDP per capita.

Figure 12. *Results of different fixed regressors using BI data.*



Note: Dashed line denotes the prior inclusion probability. lGDPpc fixed regressors include log GDP per capita. Continental fixed regressors include log GDP per capita and regional dummies. SDM fixed regressors include life expectancy, primary school enrolment, and log GDP per capita.

Figure 13. *Results of different fixed regressors using SDM data*.

Different priors

In Bayesian model averaging we had to specify priors on model probabilities and on the variance of coefficients — the ‘g prior’. This means approaching the data with different prior expectations can alter the posterior conclusions. If IQ really is robustly associated with economic growth, it should perform well under all reasonable priors. So far we have only used one set of priors. We created model probabilities based on the number of variables they included, assuming a fixed probability of the inclusion of any variable such that the expected model size was 7. We let the g prior be equal to the number of observations, calibrating our certainty in coefficient sizes to the amount of information we could supply with our Bayesian model averaging.

To test whether our results are robust to different priors, we employ all possible combinations of model priors and g-priors within the BMS package. In our results presented in Figures 14 and 15, the first part of the legend indicates what model priors were used, Random meaning model priors drawn from a beta distribution, Uniform meaning all models have the same prior, and Fixed which we have already employed. The second part of the legend provides the acronym for the g-prior used. The names, acronyms and brief explanations for these g-priors are provided in the Methodology section. Further details of these priors can be found in Feldkircher and Zeugner (2009) and Zeugner and Feldkircher (2015).

Note: Dashed line denotes the prior inclusion probability

Figure 14. *Results of different priors using BI data.*

Note: Dashed line denotes the prior inclusion probability

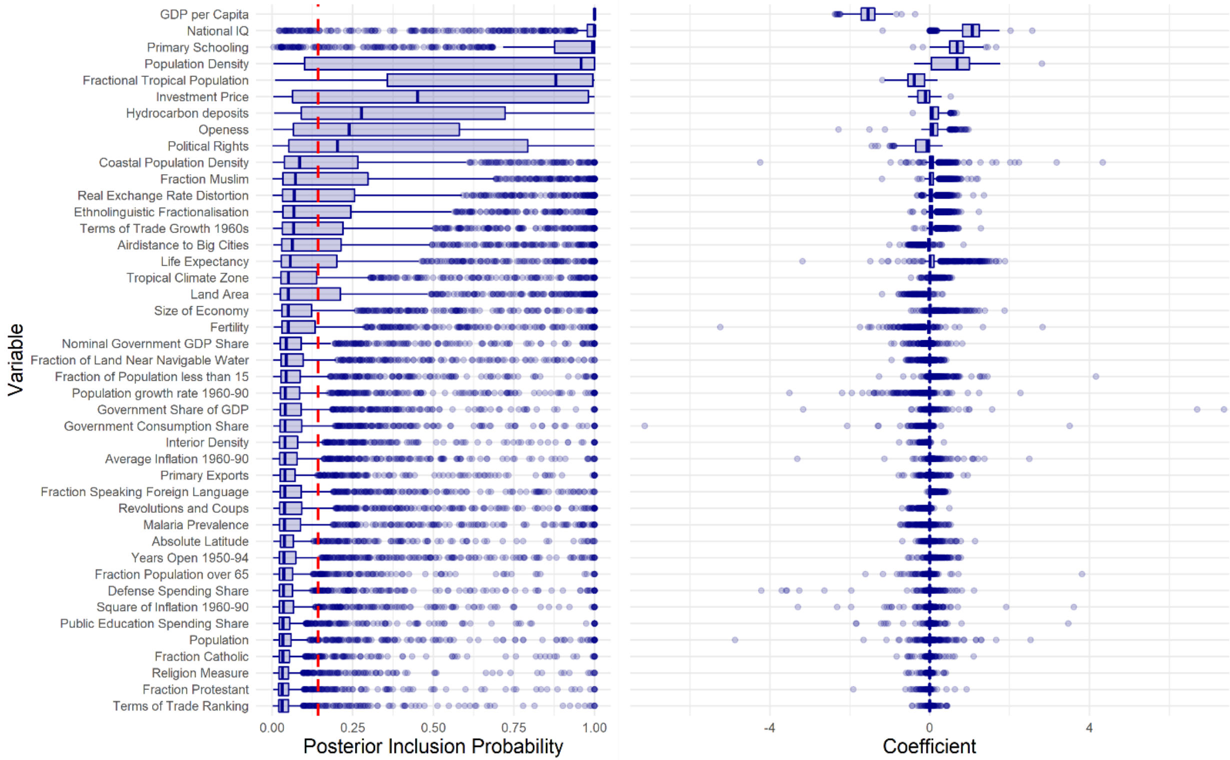
Figure 15. *Results of different priors using SDM data.*

Subsampling

In prior literature on the sensitivity of BMA in growth modelling, different data has been found to substantially change the results due to both measurement error and different sampling resulting from different datasets (Rockey & Temple, 2016). So far we have only studied whether national IQ is robust to using different Penn World Tables data and different national IQs, but we have not studied the issue of sampling in isolation. With sampling bias our results so far may be inaccurate, and even with random sampling our results may be coincidental due to outliers. To pursue this issue further we use Bayesian model averaging again by resampling our data with two methods: bootstrapping, and jackknife resampling. Thus we are performing observation sampling and then model sampling sequentially. Although resampling and weighting observations has been recommended to improve the robustness of Bayesian model averaging with economic growth (Doppelhofer & Weeks, 2011), no one has yet tried it. To perform this resampling most effectively, we use the SDM dataset of controls because its larger sample size allows us to present the results of a wider range of possible combinations of observations.

In bootstrapping we randomly resample our observations. In this process the same observation may be picked more than once for the new sample. We resample our observations 1,000 times. We then perform Bayesian model averaging upon each resample. Our initial attempt to do this ran into dummy variable traps that had homogenous values for certain variables making the regressions impossible. To solve this problem we sorted the variables by the number of unique values they had and deleted the variables with fewest unique values, one by one, until running all our regressions was feasible. This meant we had to reduce the number of variables from 68 to 43. Although we could have kept all variables and ignored failed regressions, that would be limiting the resampling, 'baking in' sampling bias.

We use the box plot shown in Figure 16 to present the distribution of the PIPs and coefficients estimated from this method.



Note: Dashed line denotes the prior inclusion probability

Figure 16. *Bootstrap sampling results using SDM data.*

Of our tested variables, GDP per capita in 1960 and national IQ had the highest median PIP of 1. Primary school enrollment in 1960 was a close competitor with a median PIP of 0.99. However, national IQ performed the best with the first quartile of its PIPs being equal to 0.98, whilst primary education's lower quartile PIP was at 0.88. National IQ had the lowest interquartile range of PIPs at 0.02. The next smallest interquartile range was for primary education at 0.12, six times larger than national IQ's PIP interquartile range. National IQ's median coefficient was the largest for all variables tested at 1.07. primary school enrolment had a coefficient of 0.69. This was the second largest coefficient, but it was still only 64% the size of national IQ's coefficient. Political Rights, a democracy index with higher values indicating greater levels of democracy (Barro, 1991), was the worst performing of all the variables with PIPs greater than priors. It should be noted its coefficient on economic growth was negative, suggesting the result may have been a fluke. Summary statistics for PIPs and coefficients of variables with a median PIP higher than the prior inclusion probability (0.14) are given in Tables 5 and 6.

Notably, PIPs range from at least 0.2 to 1 for all trialed variables. This is testament to the sensitivity of Bayesian model averaging of economic growth to sampling, suggesting that reporting one or a few BMAs with the same data or variables is not sufficient to be confident in one's results. Nonetheless, the fact national IQ is the best performing in terms of its median coefficient, median PIPs, first quartile PIP, and interquartile range of PIP in the face of this sensitivity, suggests it is the most robust predictor of economic growth.

Table 5. *Summary statistics of posterior inclusion probabilities.*

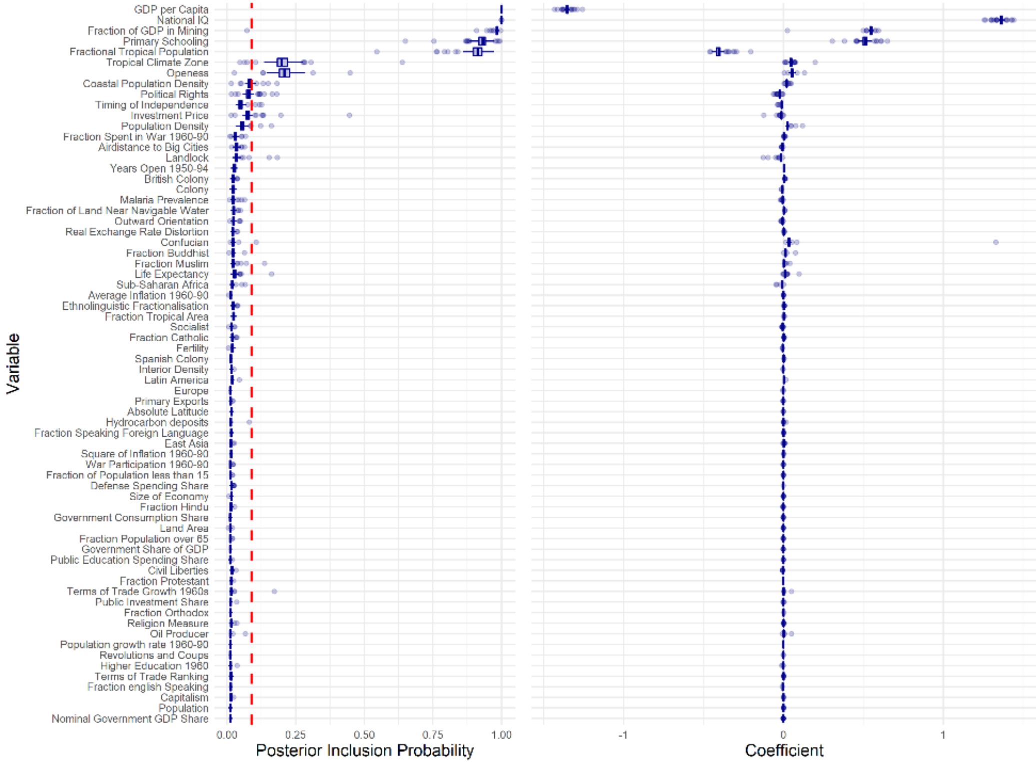
Variable	Min.	1st quartile	Median	Mean	3rd quartile	Max.	Inter-quartile range
Political rights	0.00	0.05	0.20	0.39	0.79	1.00	0.74
Openness	0.00	0.07	0.24	0.35	0.58	1.00	0.51
Hydrocarbon deposits	0.01	0.09	0.28	0.40	0.72	1.00	0.63
Investment price	0.00	0.06	0.45	0.51	0.98	1.00	0.92
% tropical population	0.01	0.36	0.88	0.69	1.00	1.00	0.64
Population density	0.01	0.10	0.96	0.65	1.00	1.00	0.90
Primary schooling	0.00	0.88	0.99	0.86	1.00	1.00	0.12
GDP per capita	1.00	1.00	1.00	1.00	1.00	1.00	0.00
National IQ	0.02	0.98	1.00	0.90	1.00	1.00	0.02

Table 6. *Summary statistics of variable coefficients.*

Variable	Min.	1st quartile	Median	Mean	3rd quartile	Max.	Inter-quartile range
Political rights	-1.45	-0.35	-0.06	-0.19	-0.01	0.32	0.34
Openness	-2.28	0.01	0.07	0.13	0.20	0.97	0.19
Hydrocarbon deposits	-0.42	0.01	0.06	0.12	0.21	0.68	0.20
Investment price	-0.55	-0.30	-0.11	-0.15	0.00	0.53	0.29
% tropical population	-1.19	-0.54	-0.39	-0.36	-0.12	0.19	0.42
Population density	-0.39	0.04	0.69	0.60	0.99	2.81	0.95
Primary schooling	-0.42	0.50	0.69	0.65	0.84	1.67	0.34
GDP per capita	-2.38	-1.71	-1.54	-1.54	-1.38	-0.36	0.33
National IQ	-1.18	0.82	1.07	0.98	1.24	2.57	0.42

Jackknife resampling, otherwise known as the leave-one-out cross validation, involves removing each observation separately and then running BMA on all the resulting subsamples. This method focuses on the effect of each one-removed observation allowing us to check if there are any influential observations skewing our estimates of national IQ's PIP or coefficient. An advantage of this is that it allows us to keep all variables from the SDM dataset employed without running into rank deficient models. However, the subsamples are more similar to the original dataset than the resamples from the bootstrapping method, making the jackknife a less rigorous check on the effects of resampling.

We find the PIP of national IQ to be 1 in all jackknife samples (Figure 17), meaning no individual observations are skewing the inclusion probability of national IQ. We find the coefficient of national IQ ranges from 1.3 to 1.4 with a median value of 1.4. Regardless of which observations are removed we find large robust coefficients for national IQ.



Note: Dashed line denotes the prior inclusion probability

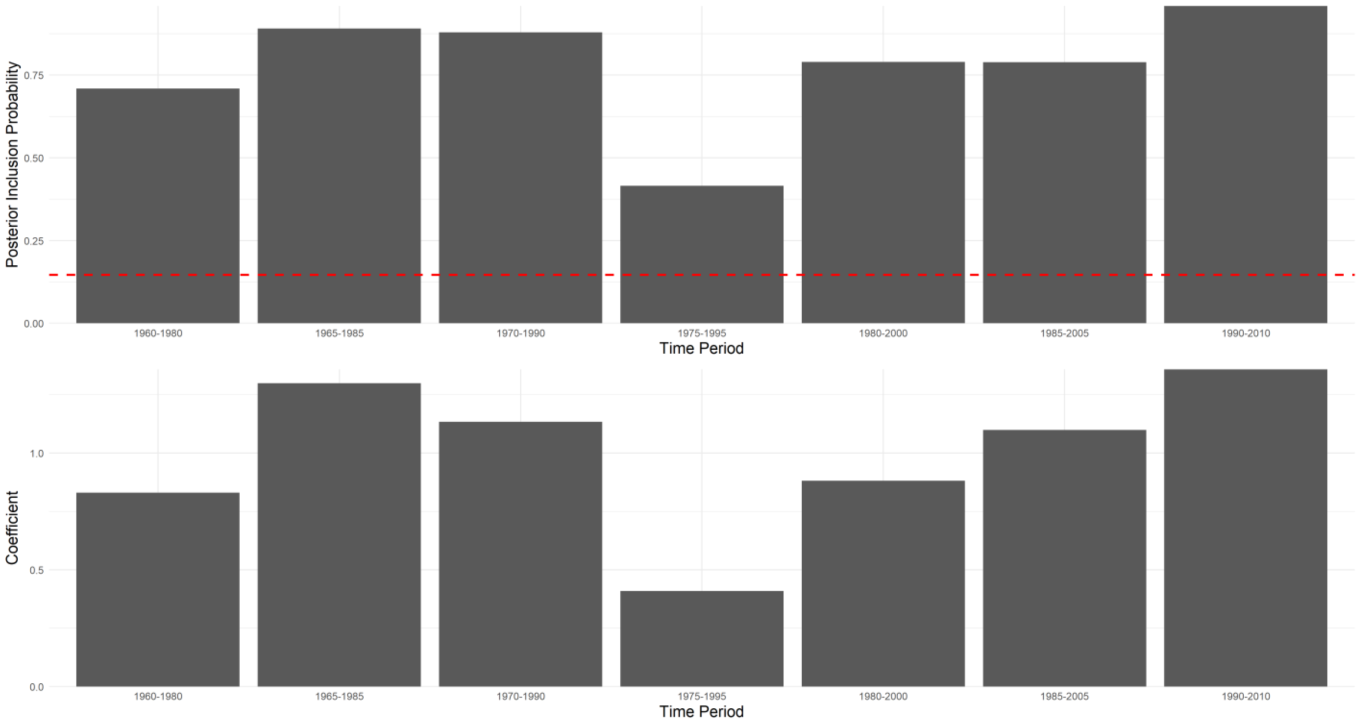
Figure 17. *Jackknife sampling results using SDM data.*

Time periods

As the BI dataset is set up for many different time periods, we re-ran our analysis on 20-year periods from 1960 to 2010. All explanatory variables, except national IQ and geographic variables, were given for the starting year of the growth period studied. National IQ had a higher PIP than prior in every single time period. Of the seven time periods studied we only found one, 1975-1995, where national IQ had a posterior inclusion probability less than 0.50. This indicates that national IQ consistently predicts economic growth, with its high performance not being the coincidental result of any particular time period. It is the best performing tested variable in all but two time periods, 1975-1995 and 1980-2000. In these time periods fertility has a higher PIP than national IQ, with a negative coefficient. This is surprising given that fertility has a lower PIP than prior in three of our subperiods and in our main 1960-2010 period. We suggest this may be coincidental due to fertility's strong negative correlation with national IQ (Meisenberg, 2009).

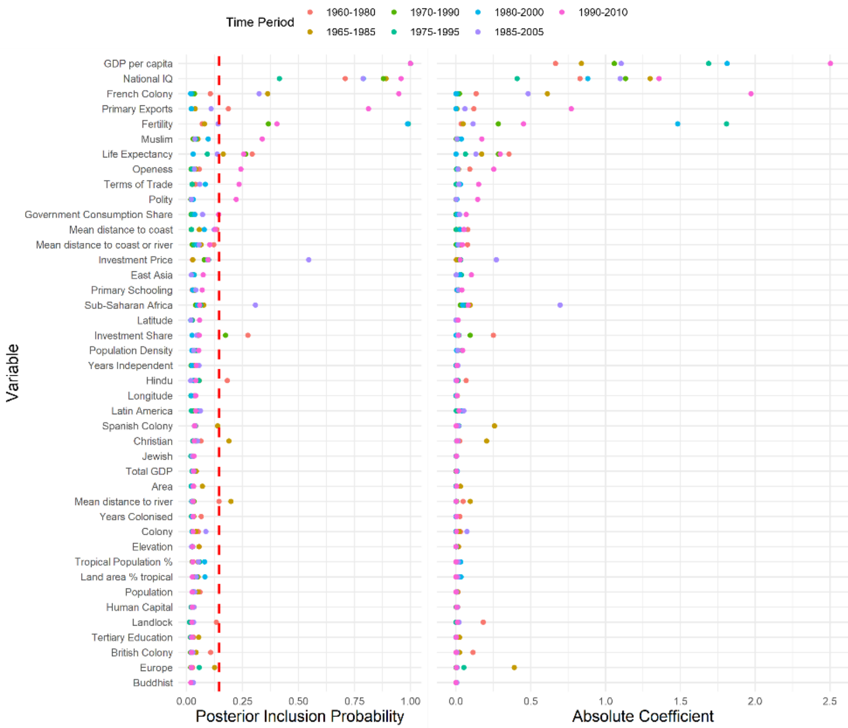
Bruns and Ioannidis (2020) were the first to test Bayesian model averaging across different time periods. They found no tested variable was robust across all time periods. They suggested this supported the view of 'robust ambiguity', that statistical modeling of economic growth is unable to identify strong explanatory variables with the exception of GDP per capita in the starting year. Their conclusion was in the title of their paper *Different Time Different Answer*. Our results contradict Bruns and Ioannidis because national IQ is supported in every one of our time periods. It is not the case that Bayesian model averaging cannot identify strong explanatory variables. Rather, economists have failed to use the variable that matters the most.

In the shorter time periods national IQ's coefficients are between 0.4 and 1.4, typically smaller than our estimate in the SDM dataset of 1.4 for the 1960-2010 time period (Figures 18 and 19). GDP per capita's coefficient is also smaller and more variable, ranging from around -0.7 to -2.5 compared to our SDM result of -1.4.



Note: Dashed line denotes the prior inclusion probability

Figure 18. *NIQ's PIPs and coefficients under different time periods using BI data.*



Note: Dashed line denotes the prior inclusion probability

Figure 19. *Results under different time periods using BI Data.*

Section 6: Causality

Whilst we have found national IQ to be extremely robust in its relationship to economic growth, the causality of this relationship can be questioned. This problem has two parts: to what extent can economic growth cause increase in national IQ scores, and to what extent do changes in national IQ scores represent real changes in 'intelligence' with the same causal effect on GDP.

The scores from psychometric tests are compiled from many different time periods mainly in the second half of the 20th century, and most of the student assessment scores are even more recent. If there is much reverse causality from growth to test scores, this could explain national IQ's strong relationship with growth.

Large increases in IQ scores in the 20th century, known as the Flynn effect, support the possibility of reverse causality. For example, IQ scores in East Asia have risen rapidly (e.g. te Nijenhuis et al., 2012). The national IQ scores we use adjust for the Flynn effect by assuming it is the same in all countries, but if Flynn effects are heterogeneous, our estimated coefficients could be upwards biased due to reverse causality. On the other hand, if Flynn effect changes in national IQ are in some way 'hollow' and have a smaller effect on GDP, then this could put a downwards bias on estimates.

An important step in disentangling the problem of causality was the study of Rindermann and Becker (2018) which found significant correlations between some lags of the Flynn effects in countries and the rate of economic growth. However, the paper only studied 27 countries and various biases could be driving the results. For example, the Flynn effect and economic growth could be confounded by a third factor, as national level fixed effects were not employed. If the Flynn effect and economic growth move in parallel, then even with lags it may be difficult to identify which factor is causing the other. This is because lagged variables do not always avoid 'simultaneity bias' (Reed, 2015). Moreover, their reported correlations do not provide us with an easily interpretable effect of Flynn effects on economic growth because they use rates of change in national IQs to predict economic growth with time periods of differing length for different observations. Furthermore, it is not clear from the results that it is not genetic changes rather than Flynn effects that are associated with economic growth.

A general problem for arguing that Flynn effects drive growth is the failure of increases in education to predict growth in fixed effects analysis (e.g. Hamilton & Monteagudo, 1998; Pritchett, 2001). A meta-analysis of quasi-experimental studies suggests years in education do increase IQ test scores (Ritchie & Tucker-Drob, 2018) meaning that if increases in IQ scores affect growth, so should education. We suggest that Flynn effects may be a hollow 'inflation' in test scores that do not affect growth. After all, whilst education increases test scores, it does not appear to affect general intelligence (Ritchie et al., 2015) or processing speed (Ritchie et al., 2013). General intelligence *g* refers to the latent factor which IQ tests try to measure (Spearman, 1904). This interpretation of Flynn effects on economic growth also concords with the Jensen effect (Rushton, 1998) whereby correlations between outcomes and IQ are strongest on more '*g*-loaded' IQ tests which are also more heritable. Many studies also support this interpretation by finding Flynn effects appear to only exist on specific IQ tests rather than on the general factor of intelligence (e.g. Jensen, 1998; Must et al., 2003; Rushton, 1998; te Nijenhuis & van der Flier, 2013; Woodley & Madison, 2013). However, some scientists have found Flynn effects represent a Jensen effect on fluid rather than crystallized measures of IQ (Colom et al., 2001). An important caveat to this line of thought is that whilst the gains from education may be hollow, other hypothetical causes of IQ increases, such as nutrition, may have "real" effects on intelligence.

An approach to remove possible reverse causality is to create national IQ scores that were created before the period of economic growth studied or very close to it, since future economic growth cannot plausibly alter past intelligence scores. This approach has been used in a few studies such as Christainsen (2020), Rindermann (2018) and Hanushek and Woessmann (2015). These studies find past test scores have the same coefficient on future growth as contemporary test scores have on past results. However, there are some limitations to this approach. The sample sizes are often much smaller. Christainsen (2020) had the largest sample size of 45 countries using this method whilst the other papers have often had substantially smaller samples. The small sample sizes may reduce the accuracy of modelling and possibly introduce bias and range restriction from missing values.

Intelligence may have a causal effect despite Flynn effects if national levels of intelligence are path dependent, which could be caused by genetics. Under such a theory societies that start more intelligent grow more and continue to perform more highly in measures of human capital, whether or not the increases in human capital measures are actually determining economic growth. This path dependence would allow us to estimate the effect of intelligence on economic growth, regardless of when intelligence is measured. For example, although Meisenberg and Woodley (2013) found the student assessment scores of low-IQ countries were catching up with high-IQ countries between 1995 and 2009, the rank order of different regions remained the same.

Strong path dependence in human capital has been shown by Baten and Juif (2014). They use age heaping measures of numeracy from 1820 and compare them with student assessment scores from the second half of the twentieth century, as used by Hanushek and Woessmann (2012). Innumerate people are less likely to be able to calculate or remember their age so they typically give rounded figures for their age, leading to ages on gravestones, censuses, and documents 'heaping' at ages that are multiples of five or ten. From the degree of 'heaping' an index for numeracy can be created. When Baten and Juif (2014) found the age heaping scores had a statistically significant relation with more recent student assessment scores, they did not report a simple correlation. Kirkegaard (2015) found Lynn's 2012 IQ scores had a correlation between 0.52 with an age heaping index in 1890 and 0.85 with an age heaping index from 1800. This approach has a few limitations because the ceiling effect of numeracy is very strong in later cohorts (i.e., there is almost no heaping), and because the set of included countries varies across cohorts.

Path dependence can be found more generally in the 'deep roots' literature on economic growth. Comin, Easterly and Gong (2010) in their paper *Was the Wealth of Nations Determined in 1000 BC* find a strong relationship between development throughout history, such as a correlation of r = .71 between 'migration adjusted technology level' in 1500 AD and log per capita income in 2002. Similar results have been found by Putterman and Weil (2010) and Spolaore and Wacziarg (2013) who diplomatically state that "The evidence suggests that economic development is affected by traits that have been transmitted across generations over the very long run."

Given strong path dependence in human capital and economic growth, it should be no surprise that contemporary IQ scores can predict past prosperity far further back than the twentieth century. For example, Lynn and Vanhanen (2012) find national IQ has a Spearman correlation with GDP per capita greater than .70 for 2003, 1870, and 1700.

Strong path dependence in human capital and GDP and correlations between them support the idea that human capital has played a large role in determining prosperity throughout history, but it is still vulnerable to collider bias with a confounding variable determining both GDP and human capital. We suggest genetic differences between populations are what determines human capital and thus GDP. Despite the controversy surrounding the issue of race and intelligence, when intelligence researchers are surveyed anonymously, 85% believe genetics plays a role in the Black-White IQ gap in the USA (Rindermann et al., 2020). Although we do not have the space in this paper to discuss racial differences in intelligence within America, if genes affect intelligence differences within the United States it is likely that they have some effect across the globe.

Genetic differences could also explain variation in the Flynn effect, with genetically smarter populations being faster to learn (eg. te Nijenhuis et al., 2012). For example, African Americans score lower than East Asians in IQ tests in

America despite similar environments. As expected, East Asian countries have seen large Flynn effects whilst sub-Saharan countries appear to have experienced none (Wicherts et al., 2010c).

Various evidence suggests the national variation in intelligence may be genetic in origin. Piffer (2015, 2019, 2020, 2021) finds educational polygenic scores, trained on predicting educational attainment in white populations, correlate with national IQs at r > .90. Nevertheless, there are concerns about the transracial validity of these polygenic scores. Economists have found genetic distance between countries is associated with various outcomes including economic growth (eg. Saha & Mishra, 2020). Whilst this literature uses genetic distance as proxy for cultural distance, assuming that ideas and technology diffuse faster across similar groups, it also has obvious implications for the possibility that genetic variation in intelligence could mediate differences in economic growth. Moreover, genetic distance correlates with national IQ (Becker & Rindermann, 2016; Kodila-Tedika & Asongu, 2016). IQs correlate with cranial size (r = .26; Pietschnig et al., 2015), and national IQs have been found to correlate with cranial size in a sample of ten countries (r = .91; Rushton, 2010), supporting a biological origin for national differences in intelligence. Furthermore, cranial capacity is substantially genetic with a heritability of around 90% in early adulthood (Batouli et al., 2014).

Quasi-experimental evidence from variation in wealth and environment also suggests genetics may be the cause of variation in national IQ. For example, countries that are or become rich due to oil wealth attain no higher IQs than their poorer genetically similar neighbors (Christainsen, 2013; Jones & Schneider, 2009). Christainsen (2013) set out to estimate environmental effects through regressing national IQ on measures of environment such as education and malnutrition. He found that regional dummy variables dominated the regression relative to environmental variables, suggesting ancestry has a much larger effect on IQ than socioeconomic environment.

A popular explanation for why nations and peoples differ in their intelligence is Cold Winters Theory (Frost, 2019; Lynn, 1987; Rushton, 1995). This theory supposes that the challenges of cold winters and the necessary preparation for the seasons is cognitively demanding such that humans are selected for intelligence further from the equator and in colder environments. The theory has repeatedly been rediscovered by scholars throughout history such as Alfred Russell Wallace (1864), Arthur Schopenhauer (2000, p. 159) and Sa'id al-Andalusi who was born in the year 1029 (Lewis & Lewis 1990, pp. 47-48). The theory fits the data as groups with higher intelligence and higher cranial capacity tend to have evolved in colder environments further from the equator (Kanazawa, 2008). The strongest cold winter correlate of national IQ appears to be skin reflectiveness suggesting UV radiation may best capture the cold winter effect (Templer & Arikawa, 2006). Furthermore, behavioral ecologists have independently found the same pattern within non-human primates (Navarrete et al., 2016), birds (e.g. Roth et al., 2010; Sol et al., 2010) and other species (Gillooly & McCoy, 2014; Jiang et al., 2015), giving the theory parsimony. In particular, the evidence from birds shows that the relationship between absolute latitude or winter temperature to brain size is only present in non-migrating birds, a prediction of the cold winters theory that is difficult to explain otherwise.

To test whether genetic or path-dependent variation in intelligence causes growth, rather than the reverse, we use instrumental variable estimation. Appropriate instrumental variables are ones that should influence national intelligence but have no other association with economic growth, allowing us to isolate the direct effect of national IQ on economic growth. In the first stage the instrumental variables model national IQ and predicted national IQs are taken. In the second stage, growth regressions are run using predicted national IQs rather than actual IQs. The predicted values of national IQ represent the effect of the instrumental variables directly on national IQ and any possible indirect effect through economic growth. Crucially, however, the predicted national IQs are unaffected by exogenous changes in GDP during the growth period studied.

We employ two further statistical tests in this approach. Firstly we use the Wu-Hausman endogeneity test. This is a test of whether the OLS and IV coefficient estimates significantly differ, indicating the existence of endogeneity in the OLS regression model. Because we expect our IV estimates to identify a causal estimate of the effect of prior cognitive ability on economic growth, a significantly different OLS estimate would indicate that endogeneity biases estimates of IQ's effect on economic growth. Furthermore, we employ the Weak Instruments Test. This is an F test comparing the second stage regression with and without the instrument. If the regression does not perform significantly better with the instrument, this suggests it is weak. To perform instrumental variable estimation we use the R package *ivreg* (Fox et al., 2021). Further details of our statistical tests can be found in the textbook *Econometric Analysis* (Greene, 1993).

We are not the first to employ instrumental variables for measures of national cognitive ability. Previously, measures of educational institutions and school quality (Hanushek & Woessmann, 2015) have been used as instrumental variables. For example, these instrumental variables include the level of private school competition, the existence of exit exams, and relative teacher pay. They found no evidence of reverse causation with the Wu-Hausman endogeneity test. However, these instrumental variables are taken from times during the growth period studied. This makes these variables unsuitable for identifying causality because educational institutions may be influenced by changes in growth or changes in intelligence. In our instrumental variable approach we only use measures taken before the period of growth to represent the deep root causes of intelligence. Christainsen (unpublished) has used regional dummies as instruments for national IQ. Given it is unpublished, we refrain from discussing the results.

We employ three instrumental variables separately. Firstly we use numeracy measures created with age heaping, which was collected from a range of samples by Joerg Baten (2015). We standardize age heaping scores across all time periods in the 19th century and then take an average. This tests whether human capital's relationship with growth is path dependent, allowing us to predict economic growth with national IQs from any time period. Our second instrumental variable is cranial capacity. This is an estimate from Beals et al. (1984) which uses a sample of skulls from 124 ethnic groups and then imputes estimated cranial capacity for all areas of the planet. From these scores David Becker (2019) estimated cranial capacities for countries, adjusting for population density and migration (See column FW in the 'NAT' tab of version 1.3.3 of the national IQ dataset at https://viewoniq.org/?page_id=9). Given the prior support for Cold Winters theory, we also use UV radiation by country adjusted for migration post-1500. This measure is obtained from Andersen et al. (2021), who estimate the average level of UV radiation nations' ancestors had in 1500. The cranial capacity and UV measures thus allow us to test whether biological factors determine the wealth of nations. A correlation matrix of our instrumental variables with Rindermann's national IQ scores is provided in Table 7.

For control variables we use log GDP per capita in the starting year. We also perform the same IV estimation with the control variables from the model with the highest posterior model probability (15%) in our Bayesian model averaging with the SDM dataset of controls. These variables are Tropical Population Percent, Primary School Enrollment in 1960, and Fraction of GDP in Mining. Initial regressions before employing IV estimation are in Table 8 and our IV estimates are in Table 9.

Table 7. *Correlation matrix of NIQ and instruments*

	Age heaping	Ancestry-adjusted UV radiation	Cranial capacity
National IQ	0.69	-0.79	0.54
Age heaping	1.00	-0.69	0.39
Ancestry-adjusted UV radiation		1.00	-0.68

Table 8. *OLS models of economic growth; * p < 0.05, ** p < 0.01, *** p < 0.001*

Variables	Model 1	Model 2
Intercept	-0.66 (0.84)	2.34* (1.05)
national IQ	0.14*** (0.01)	0.11*** (0.01)
Log GDPpc 1960	-1.06*** (0.14)	-1.31*** (0.13)
Fraction of GDP in mining		5.80*** (1.22)
Primary school enrolment		1.46** (0.44)
Tropical population percent		-0.93** (0.21)
Observations	104	94
Adjusted R^2	0.63	0.76

Table 9. *Instrumental variable models of economic growth; * p < 0.05, ** p < 0.01, *** p <0.001*

	Instrumental variable					
	Cranial capacity		Ancestry-adjusted UV radiation		Age heaping (19th century numeracy)	
Model	**1**	**2**	**3**	**4**	**5**	**6**
Intercept	-0.60 (0.89)	1.78 (2.16)	-0.91 (0.89)	1.25 (1.56)	1.29 (0.96)	3.28* (1.52)
National IQ	0.13*** (0.02)	0.13*** (0.04)	0.17*** (0.03)	0.14*** (0.03)	0.14*** (0.02)	0.11*** (0.03)
Log GDPpc 1960	-1.02*** (0.23)	-1.36*** (0.21)	-1.31*** (0.17)		-1.27*** (0.19)	-1.39*** (0.16)
Fraction of GDP in mining		5.95*** (1.32)		6.07*** (1.28)		4.98*** (1.12)
Primary school enrolment		1.28 (0.74)		1.12 (0.57)		1.31 (0.92)
Tropical population %		-0.81 (0.531)		-0.69 (0.41)		-0.86* (0.37)
Observations	104	94	101	93	75	71
Weak instruments test p-value	0.00	0.01	0.00	0.00	0.00	0.00
Wu-Hausman p-value	0.84	0.77	0.02	0.33	0.74	0.98

Our initial estimates find one IQ point increases economic growth by 0.14% without controls and 0.11% with controls, which was the same estimate our Bayesian model averaging produced. In our IV estimates all models pass the Wu-Hausman test except when ancestry-adjusted UV is employed (p < .05) without control variables in model 3. In model 3, national IQ's coefficient is 0.17 which is significantly larger than the OLS estimate of 0.11. If there is reverse causation, it causes us to underestimate the effect of national IQ. In the other instrumental variable estimations, national IQ's coefficient is similar to the OLS estimate. A likely explanation for model 3 is that ancestry-adjusted UV may not be exogenous because its geographic or environmental confounds may have an independent effect on growth. This problem might be solved by using control variables such as Tropical Population. In all regressions our instrumental variables passed the Weak Instruments test, meaning they have a robust relationship with national IQ.

The failure to reject the null in the Wu-Hausman test is suggestive of there being little reciprocal causation, but it is not proof. It merely suggests OLS estimates are close to the true causal effect size, assuming our instruments only covary with GDP growth due to their effect on intelligence.

Overall our instrumental variable estimates found results consistent with the OLS and BMA methods. We conclude that reverse causality does not play a substantial role in distorting estimated coefficients of national IQ on GDP growth.

Section 7: Limitations

The Bayesian model averaging method attains more reliable estimates by reducing researcher degrees of freedom in choosing explanatory variables. Nonetheless many researcher degrees of freedom still existed in this study such as the choice of dataset and the choice of time period. Furthermore, BMA creates more researcher degrees of freedom by requiring the specification of prior probabilities. In this study we have tried using a sample of plausible methodologies as robustness tests to see whether the sensitivity of BMA is distorting our results. We held our method constant whilst making one change to our method at a time as a robustness test. It is possible that different combinations could have produced different results. Nonetheless, given the breadth of tests performed and the consistency of national IQ to have the highest average PIP and coefficient in all the tests gives us strong reasons for supposing national IQ is the best predictor of economic growth.

A more challenging problem would be if our results were systematically biased. This could occur through sample bias and range restriction. Our data will not be missing at random as it is likely that poorer countries are less likely to have data available from the 1960s. This may undermine the power of variables competing with national IQ. It is likely that the worst socialist and authoritarian countries would not have sufficient economic statistics to be in our sample. For example, North Korea was not an observation in any of our models. Were the statistics available we might have found stronger results for economic and political freedom indexes, when in fact their posterior inclusion probabilities were systematically lower than their prior inclusion probabilities. As the world develops, more data should be released from countries allowing national IQ to be tested with large samples. However, regional levels of prosperity have consistently had a strong relationship with regional IQs (Fuerst & Kirkegaard, 2016; Lynn et al., 2018). Given that there are no sample biases for within-country studies, we should be skeptical to think that sample bias plays any substantial role in the high performance of national IQ to predict growth.

A related surprising result from Bayesian model averaging is how poorly popular theories of economic growth perform. As mentioned, popular institutional measures such as democracy and economic freedom appear to have negligible or even negative effect sizes. A possible reason for this is that our variables may be poor quality measures, for example there are various difficulties in measuring institutions (see Glaeser et al., 2004). Furthermore, many of the variables tested will change over time meaning economic growth might relate to their average values over a time period rather than their initial values. This problem should be reduced in the subperiod analysis we have performed. Yet many popular variables, such as the Polity 2 democracy index, still perform poorly in the subperiod analysis.

An important question for judging our control variables is how to interpret posterior inclusion probabilities. The very low PIPs of rival explanatory variables might suggest that only national IQ and few other variables matter. Alternatively, the rival variables might have a small but real effect on economic growth which we are unable to distinguish due to the low sample size. With only one planet of nations, of which we only have a limited sample, regression methods only have sufficient degrees of freedom to distinguish the largest effects on economic growth. Whilst it is certainly plausible that variables apart from education, IQ and natural resources do influence growth, their effects may be subtle and more suited to historical rather than statistical analysis.

Section 8: Conclusion

Of our tested variables, national IQ consistently has the largest coefficient and the largest posterior inclusion probability, suggesting it is the most robust predictor of economic growth. This replicates the finding of Jones and Schneider (2006) showing that national IQ has a high posterior inclusion probability in Bayesian model averaging. We found that IQ's effect was robust under many tests such as the use of different data, different fixed regressors, different time periods and resampling methods. Prior literature which did not use national IQ found that growth modeling led to 'robust ambiguity' without clear indications of which variables really matter. Our results contradict robust ambiguity findings because there is strong consistent support for national IQ. The methods, sample size and data quality were not insufficient to find powerful causes of the wealth of nations, rather the best explanatory variable was not being used by economists.

We also applied Bayesian model averaging to study more niche issues in the national IQ literature. We found potentially confounding variables and rival psychometric variables, but these did not explain away national IQ's relationship with economic growth. Using updated measures of smart fractions we found that only the average IQ of nations was robustly predictive of economic growth, rather than the intelligence of the 'elite' section of the intelligence distribution.

In interpreting our results, we discussed the prior literature suggesting that human capital differences between nations are deep-rooted and possibly of biological origin. To support this hypothesis we used 19th-century numeracy measures, cranial capacity, and ancestry-adjusted UV radiation as instrumental variables for national IQ. Endogeneity was only found ($p < .05$) in one of the six models. This was when ancestry-adjusted UV radiation was used as an instrument with only the logarithm of GDP per capita as a control variable. This result disappeared when additional controls were employed. We suggested the endogeneity found in this regression represented UV radiation having geographic confounds that could affect economic growth. Overall our IV methodology could not find strong evidence for reciprocal causation.

Our findings have substantial implications for government policy and the future of economic growth. The poor evidence for smart fraction theory suggests only small effects from having an intelligent elite. This weakens the case for policies, such as Paul Romer's charter cities, 'state building' and imperialism, which attempt to employ highly educated smart people from Western countries to design or run key institutions in developing countries. The finding may also suggest immigration can lower per capita GDP. If a high IQ country takes in lower IQ immigrants the new average may determine the prosperity of the society, even if the intelligence of the native elite remains the same. Moreover the finding makes the 'migration-ability paradox' (Rindermann, 2018, p. 422) worrisome. When less intelligent countries send their smartest people to intelligent countries, this can lower the average IQ of both nations. Under Smart Fraction theory the less intelligent nation might lose whilst the more intelligent nation may be relatively unaffected. However, when the average national IQ is what matters, both senders and receivers of migrants may be made worse off.

National IQ appears to be the most important factor in determining the GDP of a nation, yet it is deeply path-dependent with 19th-century numeracy measures having an effect on GDP similar to more recent student test scores. Moreover, the strong correlates of national IQ with the genetic polygenic scores and cranial capacity suggest biology determines which human capital paths nations are on. The inequality of countries may be fatalistically determined in our genes.

To find policies that can increase economic growth, economists, scientists, governments and the private sector should study and test the effectiveness of policies to increase national intelligence. Our results provide new evidence supporting Cattell's (1937a,b) calls for nations to develop strategies to increase their intelligence. For example, Cattell recommended an 'intelligence department' of the state devoted to measuring and improving national intelligence over time. With embryo selection and gene editing, humanity now has powerful and consensual tools to increase national intelligence genotypically. See Anomaly and Jones (2020) and Anomaly (2020) for a discussion of the ethics and policy implications of genetic engineering.

If genes determine GDP, we must expect future economic growth to fall. Since Charles Darwin (1871), scientists have observed dysgenics — the less intelligent having more children and doing so faster than others. See Dutton and Woodley (2018) for a review of this literature. More recently genetic data from the United States (Beauchamp, 2016), United Kingdom (Hugh-Jones & Abdellaoui, 2021) and Iceland (Kong et al., 2017) show polygenic scores for educational attainment to be declining. After accounting for unexplained variance in the educational polygenic scores in the Icelandic data, Dutton and Woodley (2018) estimated that IQ was falling by 0.8 points per decade. We can crudely extrapolate our finding that each IQ point increases GDP per capita by 7.8% to estimate the effect of dysgenics on GDP in one hundred years' time. If we could stop the current dysgenics of 0.8 points per decade, then GDP will be $e^{10\cdot 0.8\cdot 7.8\%} - 1 \approx 87\%$ higher in 2122 than under our current dysgenic trajectory. The power of genetics to determine prosperity paints a bleak picture of our future.

Online Supplement: The appendix is available at https://osf.io/4x38f/, as are Figures 1-19 from this paper.

Acknowledgements: We would like to thank the attendants of the London Conference of Intelligence 2021 for their feedback on an early presentation of this paper. We would particularly like to thank Gregory Christainsen for his discussions with the authors. All errors and omissions are our own.

References

Aghion, P., Akcigit, U., Hyytinen, A. & Toivanen, O. (2017). Living the American dream in Finland: The social mobility of inventors. https://www.college-de-france.fr/media/philippe-aghion/UPL7649133991263729056_PDF_3.pdf

Akcigit, U., Pearce, J. & Prato, M. (2020). *Tapping into talent: Coupling education and innovation policies for economic growth*. NBER Working Paper 27862. doi:10.3386/w27862.

Algan, Y. & Cahuc, P. (2010). Inherited trust and growth. *American Economic Review* 100: 2060-2092. doi:10.1257/aer.100.5.2060.

Algan, Y. & Cahuc, P. (2013). Trust and growth. *Annual Review of Economics* 5: 521-549. doi:10.1146/annurev-economics-081412-102108.

Al-Ubaydli, O., Jones, G. & Weel, J. (2016). Average player traits as predictors of cooperation in a repeated prisoner's dilemma. *Journal of Behavioral and Experimental Economics* 64: 50-60. doi:10.1016/j.socec.2015.10.005.

Andersen, T.B., Dalgaard, C.J., Skovsgaard, C.V. & Selaya, P. (2021). Historical migration and contemporary health. *Oxford Economic Papers* 73: 955-981. doi:10.1093/oep/gpaa047

Angrist, N., Djankov, S., Goldberg, P. & Patrinos, H.A. (2019). *Measuring human capital*. World Bank Policy Research Paper 8742. doi:10.2139/ssrn.3339416.

Angrist, N., Djankov, S., Goldberg, P.K. & Patrinos, H.A. (2021). Measuring human capital using global learning data. *Nature* 592: 403-408. doi: doi:10.4324/9781003014805.

Anomaly, J. (2020). *Creating Future People: The Ethics of Genetic Enhancement*. New York NY: Routledge.

Anomaly, J. & Jones, G. (2020). Cognitive enhancement and network effects: How individual prosperity depends on group traits. *Philosophia* 48: 1753-1768. doi:10.1007/s11406-020-00189-3.

Barro, R.J. (1999). Determinants of democracy. *Journal of Political Economy* 107: S158-S183.

Barro, R.J. & Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics* 32: 363-394.

Barro, R.J. & Sala-i-Martin, X. (1997). Technological diffusion, convergence, and growth. *Journal of Economic Growth* 2: 1-26. doi:10.1023/A:1009746629269.

Barro, R.J. & Sala-i-Martin, X. (2003). *Economic Growth*, 2nd ed. Cambridge, MA: MIT Press.

Baten, J. (2015). *Numeracy estimates (ABCC) by birth decade and country, both before and after 1800*. https://datasets.iisg.amsterdam/dataset.xhtml?persistentId=hdl:10622/GEQKCG.

Baten, J. & Juif, D. (2014). A story of large landowners and math skills: Inequality and human capital formation in long-run development, 1820–2000. *Journal of Comparative Economics* 42: 375-401.

Batouli, S.A.H., Trollor, J.N., Wen, W. & Sachdev, P.S. (2014). The heritability of volumes of brain structures and its relationship to age: A review of twin and family studies. *Ageing Research Reviews* 13: 1-9. doi:10.1016/j.arr.2013.10.003.

Beals, K.L., Smith, C.L., Dodd, S.M., Angel, J.L., Armstrong, E., Blumenberg, B., ... & Trinkaus, E. (1984). Brain size, cranial morphology, climate, and time machines [and comments and reply]. *Current Anthropology* 25: 301-330. doi:10.1086/203138.

Beauchamp, J.P. (2016). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences* 113: 7774-7779. doi:10.1073/pnas.1600398113.

Becker, D. & Rindermann, H. (2016). The relationship between cross-national genetic distances and IQ-differences. *Personality and Individual Differences* 98: 300-310. doi:10.1016/j.paid.2016.03.050.

Behrman, J.R., Alderman, H. & Hoddinott, J. (2004). *The challenge of hunger and malnutrition*. Copenhagen Consensus. https://www.copenhagenconsensus.com/sites/default/files/CP+-+Hunger+FINISHED.pdf.

Benhabib, J. & Spiegel, M.M. (1994). The role of human capital in economic development. Evidence from aggregate cross-country data. *Journal of Monetary Economics* 34: 143-173. doi:10.1016/0304-3932(94)90047-7.

Benhabib, J. & Spiegel, M.M. (2005). Human capital and technology diffusion. In: P. Aghion & S. Durlauf (eds.), *Handbook of Economic Growth*, pp. 935-966. New York: Elsevier.

Bishop, J.H. (1989). Is the test score decline responsible for the productivity growth decline? *American Economic Review* 79: 178-197.

Bjørnskov, C. (2017). Social trust and economic growth. In: R. Wilkes, C. Wu & E. Uslaner (eds.), *Oxford Handbook of Social and Political Trust*. doi:10.2139/ssrn.2906280

Brodeur, A., Lé, M., Sangnier, M. & Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics* 8(1): 1-32. doi:10.1257/app.20150044.

Bruns, S.B. & Ioannidis, J.P.A. (2020). Determinants of economic growth: Different time different answer? *Journal of Macroeconomics* 63: 103185. doi:10.1016/j.jmacro.2019.103185.

Bruns, S.B., Asanov, I., Bode, R., Dunger, M., Funk, C., Hassan, S.M., ... & Buenstorf, G. (2019). Reporting errors and biases in published empirical findings: Evidence from innovation research. *Research Policy* 48(9): 103796. doi:10.1016/j.respol.2019.05.005.

Caplan, B. (2018). *The Case against Education: Why the Education System Is a Waste of Time and Money*. Princeton Univ. Press. doi:10.1515/9780691201436.

Caplan, B. & Miller, S.C. (2010). Intelligence makes people think like economists: Evidence from the General Social Survey *Intelligence* 38: 636-647. doi:10.1016/j.intell.2010.09.005.

Carl, N. (2014a). Does intelligence explain the association between generalized trust and economic development? *Intelligence* 47: 83-92. doi:10.1016/j.intell.2014.08.008.

Carl, N. (2014b). Verbal intelligence is correlated with socially and economically liberal beliefs. *Intelligence* 44: 142-148. doi:10.1016/j.intell.2014.03.005.

Carl, N. (2015). Cognitive ability and political beliefs in the United States. *Personality and Individual Differences* 83: 245-248. doi:10.1016/j.paid.2015.04.029.

Carl, N. & Billari, F.C. (2014a). Generalized trust and intelligence in the United States. *PLoS ONE* 9(3): e91786. doi:10.1371/journal.pone.0091786.

Cattell, R.B. (1937a). Psychology and social progress: Mankind and destiny from the standpoint of a scientist. *Journal of Nervous and Mental Disease* 86: 108-109.

Cattell, R.B. (1937b). *The Fight for Our National Intelligence*. P.S. King & Son.

Cawley, J., Conneely, K., Heckman, J. & Vytlacil, E. (1997). Cognitive ability, wages, and meritocracy. In: B. Devlin, S.E. Fienberg, D.P. Resnick & K. Roeder (eds.), *Intelligence, Genes, and Success. Scientists Respond to The Bell Curve*, pp. 179-192. New York: Springer.

Christainsen, G.B. (2013). IQ and the wealth of nations: How much reverse causality? *Intelligence* 41: 688-698. doi:10.1016/j.intell.2013.07.020.

Christainsen, G.B. (2020). Rushton, Jensen, and the wealth of nations: Biogeography and public policy as determinants of economic growth *Mankind Quarterly* 60: 458-486.

Christainsen, G.B. (unpublished). Economic freedom and cognitive ability as determinant of economic growth: A new cross-country regression analysis of the entire pre-Covid period from 1980-2019.

Ciccone, A. & Jarociński, M. (2010). Determinants of economic growth: Will data tell? *American Economic Journal: Macroeconomics* 2(4): 222-246. doi:10.1257/mac.2.4.222.

Colom, R., Juanespinosa, M. & Garcia, L. (2001). The secular increase in test scores is a "Jensen effect". *Personality and Individual Differences* 30: 553-559. doi:10.1016/S0191-8869(00)00054-4.

Comin, D., Easterly, W. & Gong, E. (2010). Was the wealth of nations determined in 1000 BC? *American Economic Journal: Macroeconomics 2*(3): 65-97.

Cuaresma, J.C., Doppelhofer, G. & Feldkircher, M. (2014). The determinants of economic growth in European regions. *Regional Studies* 48: 44-67. doi:10.1080/00343404.2012.678824.

Cuaresma, J.C., Foster, N. & Stehrer, R. (2011). Determinants of regional economic growth by quantile. *Regional Studies* 45: 809-826. doi:10.1080/00343401003713456.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.

Doppelhofer, G. & Weeks, M. (2009). Jointness of growth determinants. *Journal of Applied Econometrics* 24(2): 209-244. doi:10.1002/jae.1046.

Doppelhofer, G. & Weeks, M. (2011). *Robust growth determinants*. CESifo Working Papers 3354.

Durlauf, S., Kourtellos, A. & Ming Tan, C. (2008). Are any growth theories robust? *Economic Journal* 118: 329-346.

Dutton, E. & Woodley of Menie, M.A. (2018). *At Our Wits' End: Why We're Becoming Less Intelligent and what It Means for the Future*. Exeter: Imprint Academic (Societas essays in political & cultural criticism).

Égert, B. (2015). Public debt, economic growth and nonlinear effects: Myth or reality? *Journal of Macroeconomics* 43: 226-238. doi:10.1016/j.jmacro.2014.11.006.

Eriș, M.N. & Ulașan, B. (2013). Trade openness and economic growth: Bayesian model averaging estimate of cross-country growth regressions. *Economic Modelling* 33: 867-883. 10.1016/j.econmod.2013.05.01410.1093/qje/qjy013

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. & Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics* 133: 1645-1692. doi:10.1093/qje/qjy013.

Feenstra, R.C., Inklaar, R. & Timmer, M.P. (2015). The next generation of the Penn World Table. *American Economic Review* 105: 3150-3182.

Feldkircher, M. & Zeugner, S. (2009). *Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in Bayesian model averaging*. IMF Working Papers 09(202). doi:10.5089/9781451873498.001.

Feldkircher, M. & Zeugner, S. (2012). The impact of data revisions on the robustness of growth determinants—a note on "determinants of economic growth: Will data tell?" *Journal of Applied Econometrics* 27: 686-694. doi:10.1002/jae.2265.

Fernández, C., Ley, E. & Steel, M.F.J. (2001a). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* 16(5): 563-576. doi:10.1002/jae.623.

Fernández, C., Ley, E. & Steel, M.F.J. (2001b). Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100(2): 381-427. doi:10.1016/S0304-4076(00)00076-2.

Fox, J., Kleiber, C. & Zeileis, A. (2021). *Instrumental-variables regression by "2SLS", "2SM", or "2SMM", with diagnostics*. Available at: https://cran.r-project.org/web/packages/ivreg/ivreg.pdf.

Frost, P. (2019). The original industrial revolution. Did cold winters select for cognitive ability? *Psych* 1(1): 166-181. doi:10.3390/psych1010012.

Fuerst, J. & Kirkegaard, E.O.W. (2016). Admixture in the Americas: Regional and national differences. *Mankind Quarterly* 56: 256-374. doi:10.46469/mq.2016.56.3.2.

Gallup, J.L., Mellinger, A.D. & Sachs, J.D. (2001). *Geography datasets*. Harvard Dataverse. doi:10.7910/DVN/SPHS5E.

Gelade, G.A. (2008). IQ, cultural values, and the technological achievement of nations. *Intelligence* 36: 711-718. doi:10.1016/j.intell.2008.04.003.

Gillooly, J.F. & McCoy, M.W. (2014). Brain size varies with temperature in vertebrates. *PeerJ* 2: e301. doi:10.7717/peerj.301.

Glaeser, E.L., La Porta, R., Lopez-de-Silanes, F. & Shleifer, A. (2004). Do institutions cause growth? *Journal of Economic Growth* 9: 271-303.

Gottfredson, L.S. (1986). Societal consequences of the g factor in employment. *Journal of Vocational Behavior* 29: 379-410. doi:10.1016/0001-8791(86)90015-1.

Gottfredson, L.S. (1997). Why g matters: The complexity of everyday life. *Intelligence* 24: 79-132. doi:10.1016/S0160-2896(97)90014-3.

Greene, W.H. (1993). *Econometric Analysis*, 2nd ed.

Grosse, S.D., Matte, T.D., Schwartz, J. & Jackson, R.J. (2002). Economic gains resulting from the reduction in children's exposure to lead in the United States. *Environmental Health Perspectives* 110: 563-569. doi:10.1289/ehp.02110563.

Hafer, R.W. & Jones, G. (2015). Are entrepreneurship and cognitive skills related? Some international evidence. *Small Business Economics* 44: 283-298. doi:10.1007/s11187-014-9596-y.

Hall, R.E. & Jones, C.I. (1999). Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics* 114: 83-116. doi:10.1162/003355399555954.

Hamilton, J.D. & Monteagudo, J. (1998). The augmented Solow model and the productivity slowdown. *Journal of Monetary Economics* 42: 495-509.

Hanushek, E. & Kim, D. (1995). *Schooling, labor force quality, and economic growth*. NBER Working Paper 5399. doi:10.3386/w5399.

Hanushek, E.A. & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth* 17: 267-321. doi:10.1007/s10887-012-9081-x.

Hanushek, E.A. & Woessmann, L. (2015). *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, MA: MIT Press.

Henrich, J.P. (2020). *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. New York: Farrar, Straus & Giroux.

Heston, A., Summers, R. & Betina, A. (2001). *Penn World Table version 6.0*. Center for International Comparisons, Univ. of Pennsylvania (CICUP).

Horvath, R. (2011). Research & development and growth: A Bayesian model averaging analysis. *Economic Modelling* 28: 2669-2673. doi:10.1016/j.econmod.2011.08.007.

Hugh-Jones, D. & Abdellaoui, A. (2021). *Natural selection in contemporary humans is linked to income and substitution effects*. University of East Anglia School of Economics Working Paper.

Hvide, H. & Oyer, P. (2018). *Dinner table human capital and entrepreneurship*. NBER Working Paper 24198. doi:10.3386/w24198.

Jensen, A.R. (1980). *Bias in Mental Testing*. New York: Free Press.

Jensen, A.R. (1998). *The g Factor: The Science of Mental Ability*. Westport, Conn: Praeger.

Jiang, A, Zhong, M.J., Xie, M., Lou, S.L., Jin, L., Robert, J. & Liao, W.B. (2015). Seasonality and age is positively related to brain size in Andrew's toad (Bufo andrewsi). *Evolutionary Biology* 42: 339-348. doi:10.1007/s11692-015-9329-4.

Johnson, S, Larson, W., Papageorgiou, C. & Subramanian, A. (2013). Is newer better? Penn World Table revisions and their impact on growth estimates. *Journal of Monetary Economics* 60: 255-274. doi:10.1016/j.jmoneco.2012.10.022.

Jones, G. (2008). Cognitive ability and technology diffusion: An empirical test. *American Economic Association Annual Meeting*. https://www.semanticscholar.org/paper/Cognitive-Ability-and-Technology-Diffusion-%3A-An-Jones-Mason/7ce646c68eb3bdad749af44a813f9a7cfde280c8.

Jones, G. (2012). Cognitive skill and technology diffusion: An empirical test. *Economic Systems* 36: 444-460. doi:10.1016/j.ecosys.2011.10.003.

Jones, G. (2013). The O-ring sector and the foolproof sector: An explanation for skill externalities. *Journal of Economic Behavior & Organization* 85: 1-10. doi:10.1016/j.jebo.2012.10.014.

Jones, G. (2016). *Hive Mind: How Your Nation's IQ Matters so Much More than Your Own*. Stanford CA: Stanford Univ. Press.

Jones, G. & Podemska-Mikluch, M. (2010). IQ in the utility function: Cognitive skills, time preference, and cross-country differences in savings rates. *SSRN Electronic Journal*. doi:10.2139/ssrn.1801566.

Jones, G. & Potrafke, N. (2014). Human capital and national institutional quality: Are TIMSS, PISA, and national average IQ robust predictors? *Intelligence* 46: 148-155. doi:10.1016/j.intell.2014.05.011.

Jones, G. & Schneider, W.J. (2006). Intelligence, human capital, and economic growth: A Bayesian averaging of classical estimates (BACE) approach. *Journal of Economic Growth* 11: 71-93. doi:10.1007/s10887-006-7407-2.

Jones, G. & Schneider, W.J. (2009). IQ in the production function: Evidence from immigrant earnings. *Economic Inquiry* 48: 743-755. doi:10.1111/j.1465-7295.2008.00206.x.

Kanazawa, S. (2008). Temperature and evolutionary novelty as forces behind the evolution of general intelligence. *Intelligence* 36: 99-108. doi:10.1016/j.intell.2007.04.001.

Kanyama, I.K. (2014). Quality of institutions: Does intelligence matter? *Intelligence* 42: 44-52. doi:10.1016/j.intell.2013.10.002.

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

Kirkegaard, E.O.W. (2015). *What exactly is age heaping and what use is it?* https://emilkirkegaard.dk/en/2015/08/what-exactly-is-age-heaping-and-what-use-is-it/.

Kirkegaard, E.O.W. (2017). *Meritocracy, not racial discrimination, explains the racial income gaps in the United States: An analysis of NLSY 1979*.

Kirkegaard, E.O.W., Bjerrekær, J.D. & Carl, N. (2017). Cognitive ability and political preferences in Denmark. *Open Quantitative Sociology & Political Science*. doi:10.26775/OQSPS.2017.02.22.

Kirkegaard, E.O.W. & Karlin, A. (2020). National intelligence is more important for explaining country well-being than time preference and other measured non-cognitive traits. *Mankind Quarterly* 61: 339-370. doi:10.46469/mq.2020.61.2.11.

Kirkegaard, E.O.W. & Carl, N. (in review). Smart fraction theory: A comprehensive re-evaluation.

Kodila-Tedika, O. & Asongu, S.A. (2016). Genetic distance and cognitive human capital: A cross-national investigation. *Journal of Bioeconomics* 18: 33-51. doi:10.1007/s10818-015-9210-7.

Kong, A., Frigge, M.L., Thorleifsson, G., Stefansson, H., Young, A.I., Zink, F., ... & Stefansson, K. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences* 114: E727-E732. doi:10.1073/pnas.1612113114.

La Griffe du Lion (2002). *The Smart Fraction Theory of IQ and the wealth of nations*. http://www.lagriffedulion.f2s.com/sft.htm.

Lerner, B. (1983). Test scores as measures of human capital and forecasting tools. *Journal of Social, Political, and Economic Studies* 8(2): 131-159.

Levine, R. & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review* 82: 942-963.

Lewis, B. & Lewis, B. (1990). *Race and Slavery in the Middle East: An Historical Enquiry*. New York: Oxford Univ. Press.

Ley, E. & Steel, M.F.J. (2007). Jointness in Bayesian variable selection with applications to growth regression. *Journal of Macroeconomics* 29: 476-493. doi:10.1016/j.jmacro.2006.12.002.

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

Lim, S.S., Updike, R.L., Kaldjian, A.S., Barber, R.M., Cowling, K., York, H., ... & Murray, C.J. (2018). Measuring human capital: A systematic analysis of 195 countries and territories, 1990–2016. *Lancet* 392: 1217-1234. doi:10.1016/S0140-6736(18)31941-X.

Lubinski, D. & Benbow, C.P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science* 1: 316-345. doi:10.1111/j.1745-6916.2006.00019.x.

Lynn, R. (1987). The intelligence of the Mongoloids: A psychometric, evolutionary and neurological theory. *Personality and Individual Differences* 8: 813-844. doi:10.1016/0191-8869(87)90135-8.

Lynn, R. & Becker, D. (2019). *The Intelligence of Nations*. London: Ulster Institute for Social Research.

Lynn, R., Fuerst, J. & Kirkegaard, E.O.W. (2018). Regional differences in intelligence in 22 countries and their economic, social and demographic correlates: A review. *Intelligence* 69: 24-36. doi:10.1016/j.intell.2018.04.004.

Lynn, R. & Meisenberg, G. (2010a). The average IQ of sub-Saharan Africans: Comments on Wicherts, Dolan, and van der Maas. *Intelligence* 38: 21-29. doi:10.1016/j.intell.2009.09.009.

Lynn, R. & Meisenberg, G. (2010b). National IQs calculated and validated for 108 nations. *Intelligence* 38: 353-360. doi:10.1016/j.intell.2010.04.007.

Lynn, R. & Vanhanen, T. (2002). *IQ and the Wealth of Nations*. Westport, Conn: Praeger.

Lynn, R. & Vanhanen, T. (2006). *IQ and Global Inequality*. Washington Summit.

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

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

Maoz, Z. & Henderson, E.A. (2013). The World Religion Dataset, 1945-2010: Logic, estimates, and trends. *International Interactions* 39(3): 265-291.

Masanjala, W.H. & Papageorgiou, C. (2008). Rough and lonely road to prosperity: A reexamination of the sources of growth in Africa using Bayesian model averaging. *Journal of Applied Econometrics* 23: 671-682. doi:10.1002/jae.1020.

Meisenberg, G. (2009). Wealth, intelligence, politics, and global fertility differentials. *Journal of Biosocial Science* 41: 519-535. doi:10.1017/S0021932009003344.

Meisenberg, G. (2012). National IQ and economic outcomes. *Personality and Individual Differences* 53: 103-107. doi:10.1016/j.paid.2011.06.022.

Meisenberg, G. & Lynn, R. (2011). Intelligence: A measure of human capital in nations. *Journal of Social, Political, and Economic Studies* 36: 421-454.

Meisenberg, G. & Woodley, M.A. (2013). Are cognitive differences between countries diminishing? Evidence from TIMSS and PISA. *Intelligence* 41: 808-816. doi:10.1016/j.intell.2013.03.009.

Mischel, W., Ebbesen, E.B. & Raskoff Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology* 21: 204-218. doi:10.1037/h0032198.

Mischel, W., Shoda, Y. & Rodriguez, M. (1989). Delay of gratification in children. *Science* 244: 933-938. doi:10.1126/science.2658056.

Murphy, R.H. & Lawson, R.A. (2018). Extending the Economic Freedom of the World Index to the Cold War era. *Cato Journal* 38: 265-284.

Must, O., Must, A. & Raudik, V. (2003). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. *Intelligence* 31: 461-471. doi:10.1016/S0160-2896(03)00013-8.

Navarrete, A.F., Reader, S.M., Street, S.E., Whalen, A. & Laland, K.N. (2016). The coevolution of innovation and technical intelligence in primates. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371: 20150186. doi:10.1098/rstb.2015.0186.

Neal, D.A. & Johnson, W.R. (1996). The role of premarket factors in Black-White wage differences. *Journal of Political Economy* 104: 869-895. doi:10.1086/262045.

Nelson, R.R. (1989). What Is private and what is public about technology? *Science, Technology, & Human Values* 14: 229-241. doi:10.1177/016224398901400302.

Nelson, R.R. & Phelps, E.S. (1966). Investment in humans, technological diffusion, and economic growth. *American Economic Review* 56: 69-75.

Nordhaus, W.D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences USA* 103: 3510-3517. doi:10.1073/pnas.0509842103.

Olsson, O. (2009). On the democratic legacy of colonialism. *Journal of Comparative Economics* 37: 534-551. doi:10.1016/j.jce.2009.08.004.

Pietschnig, J., Penke, L., Wicherts, J., Zeiler, M. & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience & Biobehavioral Reviews* 57: 411-432. doi:10.1016/j.neubiorev.2015.09.017.

Piffer, D. (2015). A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence* 53: 43-50. doi:10.1016/j.intell.2015.08.008.

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from GWAS hits: A replication of previous findings using recent data. *Psych* 1(1): 55-75. doi:10.3390/psych1010005.

Piffer, D. (2020). *Estimating cross-population LD decay and its effect on trans-ethnic polygenic scores*. Preprint, OSF. doi:10.31219/osf.io/gukhs.

Piffer, D. (2021). Divergent selection on height and cognitive ability: Evidence from Fst and polygenic scores. *OpenPsych* [Preprint]. doi:10.26775/OP.2021.04.03.

Pinkovskiy, M. & Sala-i-Martin, X. (2016). *Newer need not be better: Evaluating the Penn World Tables and the World Development Indicators using nighttime lights*. NBER Working Paper 22216. doi:10.3386/w22216.

Polity IV (2014). Polity IV project: Politial regime characteristics and transitions, 1800-2013. Available at: http://www.systemicpeace.org/inscrdata.html.

Potrafke, N. (2012). Intelligence and corruption. *Economics Letters* 114: 109-112. doi:10.1016/j.econlet.2011.09.040.

Pritchett, L. (2001). Where has all the education gone? *World Bank Economic Review* 15: 367-391.

Próchniak, M. & Witkowski, B. (2013a). Time stability of the beta convergence among EU countries: Bayesian model averaging perspective. *Economic Modelling* 30: 322-333. doi:10.1016/j.econmod.2012.08.031.

Próchniak, M. & Witkowski, B. (2013b). The analysis of the impact of regulatory environment on the pace of economic growth of the world countries according to the Bayesian model averaging. *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.2367254.

Proto, E., Rustichini, A. & Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy* 127: 1351-1390. doi:10.1086/701355.

Putterman, L. & Weil, D.N. (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *Quarterly Journal of Economics* 125: 1627-1682.

Putterman, L.G., Tyran, J.-R. & Kamei, K. (2010). Public goods and voting on formal sanction schemes: An experiment. *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.1535393.

Ram, R. (2007). IQ and economic growth: Further augmentation of Mankiw–Romer–Weil model. *Economics Letters* 94: 7-11. doi:10.1016/j.econlet.2006.05.005.

Reed, W.R. (2015). On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics and Statistics* 77: 897-905. doi:10.1111/obes.12088.

Rieger, M.O., Wang, M. & Hens, T. (2021). Universal time preference. *PLoS ONE* 16(2): e0245692. doi:10.1371/journal.pone.0245692.

Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? *Psychologische Rundschau* 57(2): 69-86. doi:10.1026/0033-3042.57.2.69.

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality* 21: 667-706. doi:10.1002/per.634.

Rindermann, H. (2008). Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence* 36: 127-142.

Rindermann, H. (2012). Intellectual classes, technological progress and economic development: The rise of cognitive capitalism. *Personality and Individual Differences* 53: 108-113. doi:10.1016/j.paid.2011.07.001.

Rindermann, H. (2018). *Cognitive Capitalism: Human Capital and the Wellbeing of Nations*, 1st ed. Cambridge Univ. Press. doi:10.1017/9781107279339.

Rindermann, H. & Becker, D. (2018). FLynn-effect and economic growth: Do national increases in intelligence lead to increases in GDP? *Intelligence* 69: 87-93. doi:10.1016/j.intell.2018.05.001.

Rindermann, H., Becker, D. & Coyle, T.R. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence* 78: 101406. doi:10.1016/j.intell.2019.101406.

Rindermann, H., Kodila-Tedika, O. & Christainsen, G. (2015). Cognitive capital, good governance, and the wealth of nations. *Intelligence* 51: 98-108. doi:10.1016/j.intell.2015.06.002.

Rindermann, H. & Thompson, J. (2011). Cognitive capitalism: The effect of cognitive ability on wealth, as mediated through scientific achievement and economic freedom. *Psychological Science* 22: 754-763. doi:10.1177/0956797611407207.

Ritchie, S.J. & Tucker-Drob, E.M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science* 29: 1358-1369. doi:10.1177/0956797618774253.

Ritchie, S.J., Bates, T.C., Der, G., Starr, J.M. & Deary, I.J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging* 28: 515-521. doi:10.1037/a0030820.

Ritchie, S.J., Bates, T.C. & Deary, I.J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology* 51: 573-582. doi:10.1037/a0038981.

Rockey, J. & Temple, J. (2016). Growth econometrics for agnostics and true believers. *European Economic Review* 81: 86-102. doi:10.1016/j.euroecorev.2015.05.010.

Roth, F. (2009). Does too much trust hamper economic growth? *Kyklos* 62: 103-128. doi:10.1111/j.1467-6435.2009.00424.x.

Roth, T.C., LaDage, L.D. & Pravosudov, V.V. (2010). Learning capabilities enhanced in harsh environments: A common garden approach. *Proceedings of the Royal Society B: Biological Sciences* 277: 3187-3193. doi:10.1098/rspb.2010.0630.

Rushton, J.P. (1995). *Race, Evolution and Behavior*. New Brunswick: Transaction.

Rushton, J.P. (1998). Secular gains in IQ not related to the g factor and inbreeding depression, unlike Black-White differences: A reply to Flynn. *Personality and Individual Differences* 27: 381-389.

Rushton, J.P. (2010). Brain size as an explanation of national differences in IQ, longevity, and other life-history variables. *Personality and Individual Differences* 48: 97-99. doi:10.1016/j.paid.2009.07.029.

Sachs, J.D., Warner, A., Åslund, A. & Fischer, S. (1995). Economic reform and the process of global integration. *Brookings Papers on Economic Activity* 1995(1): 1-118.

Saha, A.K. & Mishra, V. (2020). Genetic distance, economic growth and top income shares: Evidence from OECD countries. *Economic Modelling* 92: 37-47. doi:10.1016/j.econmod.2020.07.007.

Sala-i-Martin, X. (1997a). *I just ran four million regressions*. NBER Working Paper 6252.

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

Sala-i-Martin, X., Doppelhofer, G. & Miller, R.I. (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review* 94: 813-835. doi:10.1257/0002828042002570.

Sandefur, J. (2018). Internationally comparable mathematics scores for fourteen African countries. *Economics of Education Review* 62: 267-286.

Schopenhauer, A. (2000). *Parerga and Paralipomena: Short Philosophical Essays*, vol. 1. Oxford Univ. Press.

Schulz, J.F., Bahrami-Rad, D., Beauchamp, J.P. & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science* 366: eaau5141.

Shamosh, N.A. & Gray, J.R. (2008). Delay discounting and intelligence: A meta-analysis. *Intelligence* 36: 289-305. doi:10.1016/j.intell.2007.09.004.

Sol, D., Garcia, N., Iwaniuk, A., Davis, K., Meade, A., Boyle, W.A. & Székely, T. (2010). Evolutionary divergence in brain size between migratory and resident birds. *PLoS One* 5(3): e9617.

Spearman, C. (1904). “General Intelligence”, objectively determined and measured. *American Journal of Psychology* 15: 201-293. doi:10.2307/1412107.

Spolaore, E. & Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of Economic Literature* 51: 325-369. doi:10.1257/jel.51.2.325.

te Nijenhuis, J., Cho, S.H., Murphy, R. & Lee, K.H. (2012). The Flynn effect in Korea: Large gains. *Personality and Individual Differences* 53: 147-151.

te Nijenhuis, J. & van der Flier, H. (2013). Is the Flynn effect on g? A meta-analysis. *Intelligence* 41: 802-807. doi:10.1016/j.intell.2013.03.001.

Templer, D.I. & Arikawa, H. (2006). Temperature, skin color, per capita income, and IQ: An international perspective. *Intelligence* 34: 121-139. doi:10.1016/j.intell.2005.04.002.

Terman, L.M. (1916). The uses of intelligence tests. In: L.M. Terman, *The Measurement of Intelligence*, pp. 3-21. Boston: Houghton Mifflin. doi:10.1037/10014-001.

Thomspon, J. (2016). *Africa and the cold beauty of maths*. https://www.unz.com/jthompson/africa-and-the-cold-beauty-of-maths/.

Vivalt, E. (2019). Specification searching and significance inflation across time, methods and disciplines. *Oxford Bulletin of Economics and Statistics* 81: 797-816. doi:10.1111/obes.12289.

Wallace, A.R. (1864). *The Origin of Human Races and the Antiquity of Man Deduced from the Theory of Natural Selection*. London.

Watts, T.W., Duncan, G.J. & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science* 29: 1159-1177. doi:10.1177/0956797618761661.

Weede, E. & Kämpf, S. (2002). The impact of intelligence and institutional improvements on economic growth. *Kyklos* 55: 361-380. doi:10.1111/1467-6435.00191.

Wicherts, J.M., Dolan, C.V., Carlson, J.S. & van der Maas, H.L. (2010). Another failure to replicate Lynn's estimate of the average IQ of sub-Saharan Africans. *Learning and Individual Differences* 20: 155-157.

Wicherts, J.M., Dolan, C.V., Carlson, J.S. & van der Maas, H.L. (2010b). Raven's test performance of sub-Saharan Africans: Average performance, psychometric properties, and the Flynn effect. *Learning and Individual Differences* 20: 135-151.

Wicherts, J.M., Dolan, C.V. & van der Maas, H.L.J. (2010c). A systematic literature review of the average IQ of sub-Saharan Africans. *Intelligence* 38: 1-20. doi:10.1016/j.intell.2009.05.002.

Wolff, E.N. (2000). Human capital investment and economic growth: Exploring the cross-country evidence. *Structural Change and Economic Dynamics* 11: 433-472. doi:10.1016/S0954-349X(00)00030-8.

Woodley, M.A. & Madison, G. (2013). Establishing an association between the Flynn effect and ability differentiation. *Personality and Individual Differences* 55: 387-390. doi:10.1016/j.paid.2013.03.016.

World Bank Development Indicators (2021). Available at: https://databank.worldbank.org/source/world-development-indicators.

Yeh, Y.-H., Myerson, J. & Green, L. (2021). Delay discounting, cognitive ability, and personality: What matters? *Psychonomic Bulletin & Review* 28: 686-694. doi:10.3758/s13423-020-01777-w.

Zax, J.S. & Rees, D.I. (2002). IQ, academic performance, environment, and earnings. *Review of Economics and Statistics* 84: 600-616. doi:10.1162/003465302760556440.

Zeugner, S. & Feldkircher, M. (2015). Bayesian model averaging employing fixed and flexible priors: The BMS Package for *R*. *Journal of Statistical Software* 68(4). doi:10.18637/jss.v068.i04