J Econ Growth (2006) 11:71–93
DOI 10.1007/s10887-006-7407-2

ORIGINAL ARTICLE

Intelligence, human capital, and economic growth: A Bayesian Averaging of Classical Estimates (BACE) approach

Garett Jones · W. Joel Schneider

Abstract Human capital plays an important role in the theory of economic growth, but it has been difficult to measure this abstract concept. We survey the psychological literature on cross-cultural IQ tests and conclude that intelligence tests provide one useful measure of human capital. Using a new database of national average IQ, we show that in growth regressions that include only robust control variables, IQ is statistically significant in 99.8% of these 1330 regressions, easily passing a Bayesian model-averaging robustness test. A 1 point increase in a nation's average IQ is associated with a persistent 0.11% annual increase in GDP per capita.

Keywords Intelligence · Human capital · Economic growth

JEL Classifications O47, J24, I20

1. Introduction

The concept of human capital holds an important place in the theory of economic growth. However, the question of just how to measure a nation's stock of human capital is an unresolved issue in empirical growth research. Mankiw, Romer, and Weil (1992) kindled interest in empirically testing a Solow model that included human capital. They used a nation's rate of secondary education enrollment as their proxy for human capital. Other researchers, notably Sala-i-Martin (1997a, b) and Sala-i-Martin, Doppelhofer, and Miller (henceforth SDM) (2004), have considered primary school enrollments as one reasonable measure of human capital. And the average years of schooling measures of Barro and Lee (1993, 1994) have also received wide attention in empirical research.

G. Jones (✉)
Department of Economics and Finance, Southern Illinois University,
Edwardsville, IL 62026, USA

W. J. Schneider
Department of Psychology, Illinois State University,
Normal, IL 61790, USA

While economists commonly use education as a proxy for human capital, this widespread practice has coexisted with longstanding doubts about using school enrollments as a measure of human capital. The ability to solve problems, to think creatively, to recall facts and to reinterpret those facts in the light of changing circumstances: these are some of the key elements that economists seem to be thinking of when we think about "human capital." In describing human capital this way, we are setting aside discussion of job-specific human capital, the creation of which is analyzed in theoretical labor market models. General-purpose human capital has been the focus of growth research, and it is here that we place our focus in this paper. Fortunately for economists, psychologists spent the 20th century putting a great deal of energy into refining and improving upon one valuable technique for measuring this particular type of human capital: the intelligence test.

We use Lynn and Vanhanen's (2002) new database of IQ tests from 81 countries—tests given across the entire 20th-century—to create estimates of what Lynn and Vanhanen call "national average IQ." We use this national average IQ measure in growth regressions that also include as explanatory variables all three-variable combinations of the 21 growth variables that passed Sala-i-Martin's (1997a, b) robustness test: this implies a total of 1330 regressions. We do so in order to create a high econometric hurdle for the IQ measure. By using such robust control variables, we are able to see if the strong bivariate IQ-growth relationship ($R^2 = 43\%$) vanishes when multiple robust regressors are included in the specification.

Out of these 1330 regressions, IQ is statistically significant at the 95% level in 99.8% of the regressions, and positive in all regressions. Thus, after giving traditional growth regressors every possibility to span the same econometric space as IQ, IQ is still remarkably robust. Given these strong results, IQ easily passes the BACE (Bayesian averaging of classical estimates) robustness tests proposed by SDM (2004).

We also evaluate the explanatory power of national average IQ in growth regressions that include Sala-i-Martin's education measures. Among these 56 education-related regressions, IQ was statistically significant in every one, thus passing not only SDM's BACE robustness test, but also Leamer's (1983, 1985) more-demanding extreme bounds test. While one might expect that at least *some* linear combination of primary, secondary, and higher education measures could eliminate the statistical significance of IQ, we did not find this to be the case.

As an additional robustness check, we also show strong results for IQ when OECD countries are completely excluded from the sample. This evidence helps to address the concern that IQ tests are culturally biased in favor of people living in the developed world. And finally, we show that IQ passes Leamer's extreme bounds test at the 1% level in 455 regressions that use as controls the 18 robust growth variables from SDM (2004).

Our IQ-based results bolster the conclusions of Hanushek and Kimko (2000), who found that international mathematics and science test scores from 31 countries were strongly positively correlated with growth. Hanushek and Kimko consider the math and science scores to be indicators of "labor quality." It appears that national average IQ should likewise be considered as another robust measure of a nation's labor quality.

Changes in this index of labor quality appear to have strong effects on a nation's living standards. Results presented here, interpreted causally, imply that a 1-point increase in national average IQ will persistently raise a nation's average growth rate by an average of 0.11% per year. As is always the case in growth regressions, it is not possible to determine whether this growth effect reflects transitory catch-up growth to a higher steady state level of GDP or a permanently higher rate of steady-state growth; we discuss the theoretical and quantitative implications of both possibilities below.

The relationship between IQ and growth appears to be economically large and statistically robust, and provides more reliable results than some other popular human capital measures.

The existence of such an easily measured index of human capital should prove a boon to policymakers and to the economists who advise them. If our results prove to be as robust as they appear, then when policymakers ask the question, "What should our human capital policy seek to maximize," it appears that near the top of economists' list of responses should be the words, "National Average IQ."

We discuss below some of the policies—including improvements in early childhood nutrition, a healthier environment, expanded educational opportunities, and parental literacy—that are likely to raise this measure of human capital in developing countries.

2. Cross-cultural tests of intelligence and human capital formation

A well-constructed IQ test measures a very broad and diverse set of cognitive abilities in order to operationalize the theoretical construct of g, or general intelligence. Although it is certain that intellectual performance is multidimensional (Carroll, 1993), vast quantities of research indicate that it is the overall IQ and not the specific patterns of strengths and weaknesses that account for virtually all of an IQ test's predictive validity (Ree, Carretta, & Green, 2003).

The range of outcomes that IQ can predict with varying degrees of precision is very broad. IQ's correlation with tests of academic achievement is about .6 to .7 (Jensen, 1980, p. 319). Across all job types in the U.S. economy, the average correlation of IQ and supervisor ratings of job performance is about .3 to .5 (and the correlation is higher when job performance is measured objectively). Furthermore, IQ predicts performance better in complex occupations ($r = .56$) than simple ones ($r = .23$; (Gottfredson, 1997)). IQ correlates positively with occupational prestige, educational attainment, creativity, physical health, mental health, longevity, brain size, and nerve conduction velocity in the brain. It correlates negatively with criminal status, poverty, chronic welfare dependence, unemployment, divorce, and single-parenthood (Herrnstein & Murray, 1994; Reed & Jensen, 1992; Rushton & Rushton, 2003). The correlations for some of these outcomes are low enough that IQ has little accuracy for predicting outcomes of specific individuals and leave much variance unexplained, but it should be noted that no other psychological trait has a predictive validity even close to that of IQ for such a broad array of outcomes (Gottfredson, 1997).

General intelligence has indirect effects on worker productivity, since people with higher intelligence acquire essential job-specific knowledge more quickly and efficiently during training and on the job. General intelligence also has a direct effect on job performance when the job is inherently less trainable, such as a job that requires novel problem solving, independent decision making, and innovative adaptation (Gottfredson, 2004).

Is IQ simply an index of socioeconomic status? If so, it is difficult to explain many findings such as the fact that people with higher IQ's than their siblings (who, presumably, share the same socioeconomic status) tend to perform better in school, have higher status jobs, and earn higher incomes than their siblings (Murray, 1997, 2002).

Are IQ tests biased against women, poor people, and ethnic minorities? Responding to legitimate criticism, contemporary test developers have worked hard to develop statistical tools and common sense procedures to detect and eliminate most types of bias in IQ tests. Since the 1970s, research has repeatedly failed to demonstrate meaningful bias in terms of predictive and construct validity in major contemporary IQ tests for native-born English-speaking minority groups in the United States (Brown, Reynolds, & Whitaker, 1999; Jensen, 1980). That is, IQ predicts important outcomes equally well for these groups. If IQ tests are biased against these groups, then the criteria we use to detect such bias such as educational and occupational performance must be equally if not more biased.

Do IQ tests simply reflect the biases of Western Civilization? If they do, it is difficult to explain why East Asians (even from poor countries such as China) slightly outperform Europeans on IQ tests designed by Western scientists (Lynn, 1987). Although within a society it is useful to measure verbal knowledge and reasoning, language and cultural differences make such measurements problematic for purposes of cross-cultural research. Psychologists have developed many types of tests that measure reasoning ability using visual figures and patterns that minimize the effects of language and cultural differences. These tests, so-called "Culture-Fair" or "Culture-Reduced" intelligence tests, have roughly the same validity coefficients in predicting important outcomes as more culturally loaded tests (Court,1991). An important advantage of using "Culture-Reduced" non-verbal intelligence tests as measures of human capital instead of using measures of academic skills is that more precise estimates of human capital can be made in nations with high rates of illiteracy. Furthermore, a well-constructed "Culture-Reduced" IQ test measures a much broader array of potentially important abilities than does an academic achievement test.

Although it is certain that general intelligence is strongly influenced by genetic factors, it is equally clear that there are many environmental effects on IQ and the brain (Sternberg & Grigorenko, 2001). It is thus reasonable that a society could make changes to maximize the cognitive abilities of its population. Indeed, it appears that many societies have been doing so successfully for several decades. Among the group of countries for which there exist time-series data on that nation's average IQ, measured IQ's appear to have risen an average of two to three points per decade during the second half of the twentieth century, a phenomenon known as the *Flynn Effect*, after Flynn (1987).

The fact that some nations' performance on IQ tests have been rising faster than others has possible Solow-like convergence implications. For example, in Kenya, average IQ scores increased by a rate almost three times greater than the average rate of increase in industrialized countries over the 14-year period of 1984–1998. The factors positively associated with IQ gains appeared to be parental literacy, shrinking family size, and improved childhood nutrition and health (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003).

The possible impact of education on IQ should be noted: While estimates of education's impact on IQ vary, Winship and Korenman (1997) survey the literature of U.S. and Scandanavian studies and perform their own analyses of the U.S. National Longitudinal Surveys of Youth. An additional year of education is estimated to raise IQ by anywhere from 1.0 to 4.2 IQ points according to their literature survey. In their own regressions, point estimates range from 1.8 to 2.7 depending on the specification. They note that the true effects could be higher or lower, if the NLSY's variables contain measurement error. Neal and Johnson (1996), using quarter-of-birth as an instrument for exogenous education, find estimates toward the upper end of these ranges.

Armor (2003) reviews evidence from data mainly collected in the United States that suggests that by simultaneously optimizing environmental factors that impact intelligence, IQ can be increased by as much as 10 points (.67 standard deviations). Most of Armor's recommendations fall into two categories: increasing children's knowledge by devoting more resources to each child and promoting optimal brain development though better nutrition, healthcare, and public health policies for young children and their mothers. Specific factors that seem to have small but significant effects on intelligence include: reduced exposure to environmental toxins such as lead and mercury, reduced exposure to molds, parasites and pathogens that cause common childhood diseases, breastfeeding, nutritional supplementation, parenting behaviors that promote a cognitively stimulating and emotionally supportive home environment, increased parental education, greater family income, smaller family size, stable family structure, two-parent families, and reduced teenage pregnancy rates (Armor, 2003, pp. 92–98).

It is not clear how generalizable these estimates are to developing nations but it is reasonable to assume that optimizing environmental factors in initially more unfavorable conditions can produce even greater improvements in IQ. It appears that with the exception of increased education, most of the relevant environmental factors primarily exert their influence before the child is three years old.

This brief review of the validity of IQ tests and of IQ's malleability across generations only scratches the surface of a voluminous literature that is virtually univocal in its support of the utility and validity of IQ tests (Neisser et al., 1996). For a non-technical explanation of intelligence the reader is referred to Seligman (1992); for a more technical summary of the literature on the physiological, genetic, and behavioral observations supporting the existence of *g*, a general factor of intelligence, Jensen (1998) is highly recommended. In addition, Gottfredson (1997) has written a comprehensive yet accessible review of the occupational correlates of IQ.

3. Data

As noted above, we borrow much of our data from Sala-i-Martin's "I Just Ran Two Million Regressions." His dataset—available at his website, www.columbia.edu/~x23—was chosen because it is widely known and widely used. Further, given the fact that we introduce one entirely new variable into the empirical growth literature, it would have been cumbersome to explain and justify the details of an entirely untested set of growth data. One especially valuable feature of Sala-i-Martin's dataset is that he made every effort to use values estimated at the beginning of the period (1960) to limit the endogeneity problems that are endemic to empirical growth research. The names of the variables we use—the 21 variables that passed his robustness test, the three variables used in all regressions, and his education measures—are included in Table 1. For further information on the Sala-i-Martin dataset, as well as for a methodological critique of Sala-i-Martin's methodology, Hoover and Perez (2004) is invaluable.

Our IQ data come from Lynn and Vanhanen (2002, henceforth LV). Lynn, a psychologist, and Vanhanen, a political scientist, assembled a database of IQ tests from 81 different countries. These scores were derived from a variety of different types of intelligence tests given between the 1910s and the 1990s, using "Culture-Fair" or "Culture-Reduced" tests where possible. The majority of scores come from the 1950s through the 1990s.

According to LV, the world's average IQ (not weighted by population) was 88.2 and the standard deviation of world IQ was 11.4. As a point of reference, note that the average British IQ is defined as equal to 100, and the population standard deviation within Great Britain is 15. The reader who is interested in further detail regarding the database is encouraged to consult Appendix 1 of Lynn and Vanhanen (2002); LV's underlying IQ data from 163 tests are available online at Sailer (2004).

Lynn and Vanhanen, in their original work, reported the results of a univariate regression of the level of a nation's GDP per capita in 1998 (not the more common log-level) on IQ and a constant for 81 countries, and report that one additional IQ point is associated with a $519 increase in 1998 GDP per capita; this regression had an R^2 of 53%.

LV also performed some simple multivariate analyses using measures of political and economic freedom as additional explanatory variables; however, these multivariate analyses

Table 1 Variables from Sala-i-Martin (1997a,b)

Twenty-one variables passing Sala-i-Martin's "Two Million Regressions" test, in rank order:
1. Equipment Investment +
2. Number of Years Open Economy +
3. Fraction Confucian +
4. Rule of Law +
5. Fraction Muslim +
6. Political Rights +
7. Latin America Dummy −
8. Sub-Saharan Africa Dummy −
9. Civil Liberties +
10. Revolutions and Coups −
11. Fraction of GDP in Mining +
12. Std. Dev. of Black Market Premium −
13. Fraction of GDP in Primary Exports in 1970 −
14. Degree of Capitalism +
15. War Dummy −
16. Non-Equipment Investment +
17. Absolute Latitude +
18. Exchange Rate Distortions −
19. Fraction Protestant −
20. Fraction Buddhist +
21. Fraction Catholic −
Variables included in all Sala-i-Martin regressions
Log (GDP per capita, 1960) −
Rate of Primary School Enrollment, 1960 +
Life Expectancy, 1960 +
Other education measures
Rate of Secondary School Enrollment −
Rate of Higher Education Enrollment −
Average Years of Primary Education in Total Population −
Average Years of Secondary Education in Total Population +
Average Years of Higher Education in Total Population −
Average Years of Overall Education in Total Population −
Average Years of Overall Education in Total Population* log(GDP per capita, 1960) −
Percent of GDP Spent on Education +

Note: + and − signs indicate whether more of that value appears to be good or bad for economic growth in the 1960–1992 period, according to Table 1 of Sala-i-Martin (1997b)
Source: Sala-i-Martin (1997a, b)

used interpolated IQ data for 104 additional countries, to create an artificial "dataset" of 185 countries.

These interpolations were often based on methods that we do not endorse (e.g., assuming that members of a specific racial group have the same average IQ regardless of the country they live in), and therefore we exclude all of LV's interpolated data from our study. Two of LV's 81 original observations (for Peru and Columbia) also relied heavily on a form of interpolation, and so we exclude these observations from our dataset.

We discard another nine of Lynn and Vanhanen's 79 non-interpolated observations, either because the sample size in the particular country was not stated or was less than 100, or because the IQ estimate relied solely on the scores of emigrants. This leaves us with 70 usable observations. Table 2 provides a complete list of these 70 estimated national average IQ's by country.

Table 2 Estimated National Average IQ

	IQ		IQ
Argentina*	96	Kenya*	72
Australia*	98	Korea, South*	106
Austria*	102	Lebanon	86
Barbados	78	Malaysia*	92
Belgium*	100	Marshall Islands	84
Brazil*	87	Mexico*	87
Bulgaria	93	Morocco*	85
Canada*	97	Nepal*	78
China	100	Netherlands*	102
Congo (Brazzaville)*	73	New Zealand*	100
Congo (Zaire)*	65	Nigeria	67
Croatia	90	Norway*	98
Cuba	85	Philippines*	86
Czech Republic	97	Poland	99
Denmark*	98	Portugal*	95
Ecuador*	80	Puerto Rico	84
Egypt*	83	Qatar	78
Fiji*	84	Romania	94
Finland*	97	Samoa (Western)	87
France*	98	Singapore*	103
Germany*	102	Slovakia	96
Ghana*	71	Slovenia	95
Greece*	92	South Africa*	72
Guatemala*	79	Spain*	97
Guinea	66	Sudan	72
Hong Kong*	107	Sweden*	101
Hungary	99	Switzerland*	101
India*	81	Taiwan*	104
Iran*	84	Tanzania*	72
Iraq*	87	Turkey*	90
Ireland*	93	Uganda*	73
Israel*	94	United Kingdom*	100
Italy*	102	United States*	98
Jamaica*	72	Uruguay*	96
Japan*	105	Zambia*	77

Note: Asterisk indicates inclusion in regression results reported in Tables 3 and 4
Source: Lynn and Vanhanen (2002)

Because some of the countries included in the LV dataset are not included in Sala-i-Martin's dataset, our regressions include a maximum of 51 countries. The mean IQ in this dataset is 90.2 and the standard deviation of IQ is 11.4. As noted above, this sample of 51 is substantially larger than Hanushek and Kimko (2000), who relied upon math and science tests from 31 countries.

The national IQ estimate used in our research is the same used by LV: an average of all same-country IQ studies. However, for 36 of LVs 81 countries, LV rely on just one IQ study to estimate that nation's average IQ. This raises the question of whether one study is enough to estimate a nation's average IQ. LV answer this question by analyzing the distribution of IQ scores across various studies of the same country. In these cases, the within-country correlation between each study's average IQ scores for that country is 0.939. This high intra-country correlation across studies provides some confidence that one study alone provides a reasonable estimate of a nation's average IQ. We look forward to reassessing our results as more comprehensive databases of world IQ estimates become available.

Two published studies have used LV's data in growth regressions: Weede and Kampf (2002) and Volken (2003). However, both studies used LV's interpolated data as well as the authentic data, which may distort their results. Weede and Kampf report the results of 14 regressions, some of which include the Barro-Lee (1996) education measures along with other education measures. They find that national IQ has a large and statistically significant relationship with growth, even controlling for education measures. Volken, using a similar dataset focusing on education, reports results from 10 regressions, and finds that the relationship between IQ and growth becomes unstable once certain education variables are included. We believe that these inconsistent results probably reflect the decision to use LV's interpolated data.

We follow the practice of LV, who assume a Flynn effect of two or three points per decade, depending on which exam was given. For example, the Iranian average IQ, based on a 1957 test, was estimated as equal to 80 when compared to the a similar British test given in 1979. Because of the Flynn effect, LV assume that Iranian IQ's have risen by an average of two points per decade since 1957, so Iran's average estimated IQ is inflated to 84 in LV's dataset. This adjustment, while not ideal, follows the best practice of the psychological profession. Further, it allows us to treat all national IQ scores listed in Table 2 as being in what economists might think of as "Real 1979 IQ."

To give a first impression of the data, Figure 1 is a simple plot of IQ against real GDP per capita in 1992, measured in Summers–Heston purchasing-power adjusted dollars. The R^2 from a bivariate regression of log 1992 GDP per capita on the LV IQ measure is 79%, with one IQ point associated with an 8.7% rise in living standards.

We also note that if we calculate the R^2 from as bivariate regression of the economic growth rate on each of the explanatory variables used in this study (using data from just the 70 countries included in our IQ database), the R^2 values range from a low of 0.5% (for percent Catholic) through a high of 43% (for national average IQ). Thus IQ levels appear to have a particularly strong bivariate relationship with both the level and the growth rate of output per capita. In this bivariate setting, one IQ point is associated with a 0.10% higher growth rate over the period.

For comparison, we note that the five variables with the next highest bivariate R^2 relationships with economic growth are years the economy was open to trade (39%), percent of GDP devoted to equipment investment (34%), percent Confucian (34%), percent Buddhist

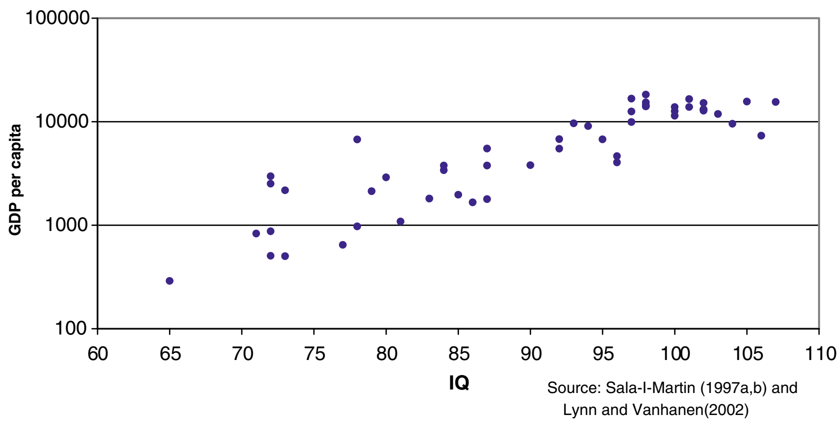

Fig. 1 IQ and 1992 GDP per capita (Summers-Heston PPP$)

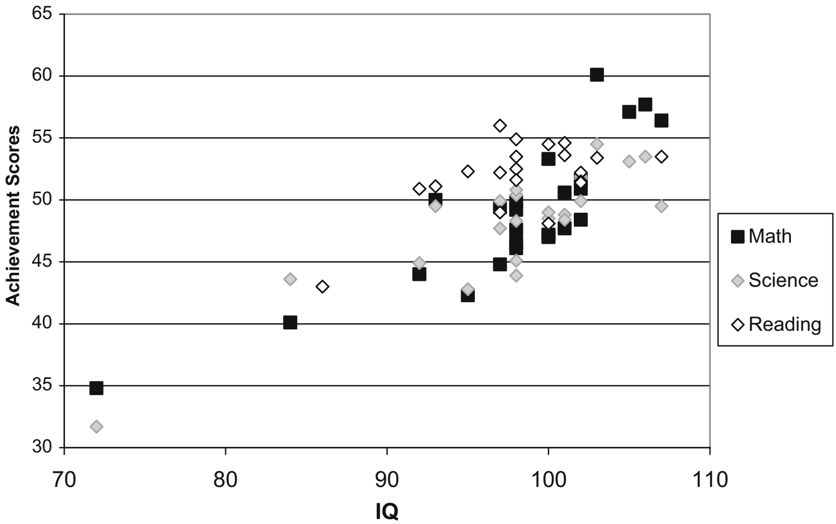

Fig. 2 IQ and Barro–Lee (1993) achievement measures

(32%), and the percent of GDP devoted to primary exports in 1970 (31%). All correlations are positive with the exception of percent Catholic and percent primary exports.

In summary, while LV's dataset has its problems, theirs is the first comprehensive attempt to assemble studies of IQ from around the world in such a way as to allow direct, international comparisons. We hope that the results we present will encourage others to delve more deeply into these intelligence tests. A comprehensive time-series database of such tests would be a natural next step.

To give an overall impression of how these IQ data compare with test scores used by other growth economists, Figure 2 shows the relationship between these IQ measures and the Barro–Lee (1996) national educational achievement scores for math, science, and reading from 23 countries. Univarate regressions of math, science, and reading achievement scores on IQ yield R^2s of 70%, 75%, and 34%, respectively. Figure 3 compares IQ to Hanushek and Kimko's (2000) two indices of national labor quality (which they denote QL1 and QL2), based upon math and science examinations. Hanushek and Kim (1995) describe how these labor quality measures were constructed. Our IQ observations overlap with 27 of Hanushek and Kimko's 31 observations. Univarate regressions of QL1 and QL2 scores on IQ yield R^2s of 83% and 74%, respectively.

Note that the Barro–Lee math and science scores have a particularly strong relationship with national average IQ, and both of Hanushek and Kimko's measures correlate positively with IQ. These correlations provide some reason to believe that all of the exams measure a similar set of mental abilities, however imperfectly. This strong positive correlation would come as no surprise to cognitive psychologists, who, as noted above, have found that outcomes on tests of mental ability invariably positively correlate with each other, with the correlation strongest when the test performance relies on what psychologists refer to as "general intelligence." A key reason to prefer IQ scores is because such scores are much more widely available than other standardized test scores, and because cognitive psychologists have a rich empirical and clinical literature devoted to making scores from different IQ tests comparable.

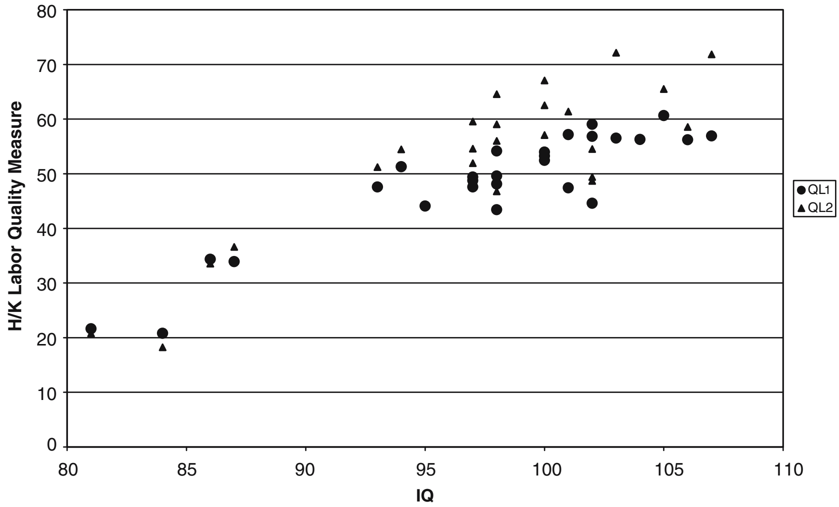

Fig. 3 IQ and Hanushek and Kimko (2000) labor quality estimates
Note: QL1 and QL2 are indices of labor quality used in Hanushek and Kimko (2000) and developed in Hanushek and Kim (1995).

Finally, note that to the extent that our IQ data mismeasure the actual human capital of the population, and to the extent that such mismeasurements come in the form of independently distributed error terms, the resulting errors in variables will generally tend to bias our coefficient estimates downward. Therefore, if IQ is a *Mismeasure Of Man*, in Gould's formulation (1981), then our estimates of IQ's impact on growth may well be too small.

4. Methodology

Since so many variables could plausibly have an impact on economic growth, the 1990s witnessed a flood of articles that each introduced and tested the statistical significance of a "new" variable, such as a nation's land mass, percent Protestant, or percent of GDP devoted to equipment investment. Many variables were found to have a positive relationship with growth, but economists were skeptical about whether any causal relationship was underlying the regressions results, and were also concerned that perhaps the statistical significance was an artifact of which control variables happened to be included in the regression—in short, economists faced model uncertainty. A rich literature developed that attempted to create robust growth regressions (*inter alia*, Bleaney and Nishiyama 2002; Brock and Durlauf, 2001; Levine and Renelt, 1992; Sala-i-Martin, 1997a,b; Sala-i-Martin, Doppelhofer, and Miller, 2004; Temple, 2000).

To address these concerns, our methodology is in the spirit of SDM (2004) which represents a refinement of the methods used in Sala-i-Martin (1997a b). SDM's approach, which they denote as "Bayesian Averaging of Classical Estimates," or "BACE" for short, starts with a sizable set of variables plausibly related to growth, and then runs (a random sample of) every single possible regression with that set of variables. SDM then present summary statistics designed to give the reader of sense of how robustly a particular variable was correlated with growth.

The SDM approach is a form of Bayesian model averaging, an approach that is increasing in popularity among empirical economists who are concerned with model uncertainty (*inter* *alia*, Wright, 2003; Koop and Potter, 2003; Milani, 2004). Since model uncertainty is the *sine qua non* of empirical growth research, it seems especially appropriate for our purposes.

In Bayesian model averaging, the researcher faces a large set of possibly true models, and affixes a "prior probability of being true" to each model. She then runs regressions for each model and then uses some function of the each regression's summary statistics to update the probability of each model's truthfulness. If one model (or class of models) performs better then expected, then the researcher increases her belief in that model (or class of models).

We follow SDM's Bayesian model averaging approach—which they refer to as Bayesian Averaging of Classical Estimates, or BACE—in order to demonstrate that the relationship between IQ and economic growth is not a mere coincidence, and that it is a relationship as strong as that between such canonical growth variables as equipment investment or number of years the economy has been open to trade. In so doing, we implicitly run a stricter Leamer-style (1983, 1985) "extreme bounds test" on IQ, the results of which we also report. Leamer's test rejects a variable if it is *ever* statistically insignificant at conventional levels in *any* regression—a very strict criterion indeed, one that only one of Sala-i-Martin's growth variables, percent Confucian, is able to pass.

The key question we want to answer in this section is whether IQ has a robust statistical relationship with a nation's average growth rate from 1960 to 1992, the time period studied by Sala-i-Martin (1997a, b). We run four sets of regressions, all of which use the average growth rate of per capita GDP from 1960 to 1992 as the dependent variable. Following Sala-i-Martin (1997a, b), each regression includes a total of seven explanatory variables: log per capita GDP in 1960, percent of the age-relevant population enrolled in primary school in 1960, life expectancy in 1960, the nation's estimated average IQ (the variable of interest), and three additional control variables.

We include the three "fixed" variables partly in order to hew as closely as possible to Sala-i-Martin's original methodology. This is because we want to tie our hands as tightly as possible in the regression specification process in order to avoid the perception of data mining or selective reporting of results. We note that the GDP per capita and primary school enrollment variables passed SDM's test, which helps to justify including these variables in all regressions.

The three additional control variables are drawn from one of two sets: the 21 variables that passed Sala-i-Martin's robustness test, or the eight measures of human capital included in Sala-i-Martin's original dataset. Note that none of these eight education measures passed Sala-i-Martin's robustness test. The aforementioned primary school enrollment variable was his best-performing human capital variable, and primary enrollment later proved to be an extremely robust growth variable in SDM (2004).

We run a separate IQ-inclusive regression for every possible combination of these variables. This implies that there are $1330 = 21!/(18!3!)$ regressions in the 'top 21' set, and $56 = (8!/(5!3!))$ regressions in the 'education' set. As noted in the introduction, we rerun all results excluding the OECD countries, in order to address the concern that IQ tests may be biased in favor of the world's developed Western countries.

To summarize our regression results for IQ, we use SDM's aforementioned BACE framework. Essentially, BACE creates a weighted average of the regression coefficients (and their respective variances) across all regressions. The weighting—defined formally below—is a function of each regression's error sum of squares, the number of observations, and the number of regressors being implicitly tested.

BACE also creates a probability measure that reflects whether looking at the data makes us more rather than less likely to include a given explanatory variable. In order to calculate this "posterior inclusion probability," we need to know whether seven-variable models that include IQ perform better (in a sense defined by SDM) than seven-variable models that exclude IQ. Therefore, in the "top 21" case, we run an additional $5,985 = 21!/(17!4!)$ regressions (for the case of education regressors, we run an additional 70).

Following the BACE framework, we employ a diffuse, non-informative prior distribution for each possible regression model. We thereby implicitly assume that each possible regression model *j* has the same prior probability, $P(M_j) = P(M)$, of being the true model (below, we relax this assumption in order to see what happens if IQ-inclusive models are given lower prior probabilities). We then update that prior distribution to arrive at a posterior probability according to the following weighted-likelihood formula,[1] comparable to Eq. (7) of SDM:

$$P(M_j|y) = \frac{P(M)T^{-k/2}SSE_j^{-T/k}}{\sum_{i=1}^{n} P(M)T^{-k/2}SSE_i^{-T/k}} \tag{1}$$

Here, *y* is the observed data, *T* is the sample size of the regression (varying from regression to regression, due to data limitations), *k* the number of explanatory variables that are being rotated through the regression (always equal to four—the three explicitly rotating variables plus IQ), SSE_j is the sum of squared errors from the linear regression of model *j*, and *n* is the total number of regressions considered. We experimented with setting *T* equal to a fixed number—the maximum number of possible observations—and with setting *k* equal to 7—the total number of regressors—and found that this did not substantively change our results.

Our approach differs from SDM in only one important respect: We consider only models of a fixed size: seven regressors, four of which are essentially fixed. SDM considered models with both more and fewer regressors, but found that model size only influenced the results for marginal growth variables. As we shall see below, IQ does not fall into that category.

We use (1) in two different ways, just as SDM do. First, we calculate (1) for the case where *n* equals the total number of possible regressions. When we consider all regressions that include Sala-i-Martin's top 21 plus IQ, we have $7,315 = 22!/(18!4!)$ regressions. Therefore, $P(M) = 1/7315$ in this case. A comparable calculation is made for the case of education controls.

Note that $P(M)$ is the probability that a given *model* is true—not that a given *variable* demands inclusion in the model. Since IQ is included in 1330 of these regressions, the prior inclusion probability for IQ is assumed to equal $1330/7315 \approx 18.2\%$. For the education variables, the prior inclusion probability for IQ is $56/126 = 44.4\%$. Below, we estimate how sensitive our results are to this prior inclusion probability.

The *SSE* and *T* from all 7315 regressions are calculated to compute the denominator for (1). We then sum (1) over all 1330 models that include IQ, and the result is the posterior inclusion probability for IQ, analogous to column (1) from Table 2 of SDM. This number is, by construction, between 0 and 1; if it is greater than the prior inclusion probability of 18%, then this indicates that models that include IQ fit better (in a weighted SSE sense) than models that exclude IQ. As SDM note (p. 823), "For [such] variables, our belief that they belong in the regression is strengthened once we see the data and we call these variables 'significant.'"

But other summary statistics may prove more informative—as well as more intuitive—than this posterior inclusion probability. Accordingly, we follow SDM in calculating weighted averages of the IQ coefficient estimates, β_{IQ}, in each set of regressions. Consider the "top 21" case: Again the weighting is set according to (1), but as in SDM (2004), now *n* is only

[1] As SDM note, Leamer (1978, p.112) derives this weighting, and Zellner (1971) provides related theoretical justification. This weighting and the Schwarz (1978) model selection criteria have common theoretical roots.

equal to 1330—so this is the expected value of β_{IQ} conditional on inclusion. In other words:

$$\bar{\beta}_{\mathrm{IQ}} = \sum_{j=1}^{1330} P(M_j|y)\beta_{\mathrm{IQ},j}$$

The variance of $\bar{\beta}_{\mathrm{IQ}}$ is also calculated as in SDM:

$$\bar{\sigma}_{\mathrm{IQ}}^2 = \sum_{j=1}^{1330} P(M_j|y)[\sigma_{\mathrm{IQ},j}^2 + (\beta_{\mathrm{IQ},j} - \bar{\beta}_{\mathrm{IQ}})^2]$$

Therefore, our estimate of the variance of the coefficient estimate takes into account both the variance of each coefficient estimate as well as the variance of estimates between regressions. The square root of these variances are the BACE-weighted standard errors we report in Table 3.

One is then tempted to take the ratio of the BACE-weighted mean coefficient and the BACE-weighted mean variance to create t-statistics, a temptation in which we will intermittently—and cautiously—indulge. While the associated standard errors are not distributed according to the usual t-distribution, SDM note that in most cases, "having a ratio of posterior conditional mean to standard deviation around two in absolute value indicates an approximate 95-percent Bayesian coverage region that excludes zero." These "pseudo-t statistics," as we will refer to them, are useful in conveying how high the signal-to-noise ratio is when it comes to IQ.

In the interest of transparency, we also report unweighted averages of the coefficients and standard errors, and find little difference between unweighted and BACE-weighted results in most cases. Another statistic we report in the table is the minimum value of the lower end of the 95% confidence interval ($\beta_{\mathrm{IQ}} - 1.96^{*}\sigma_{\mathrm{IQ}}$) across all regressions in that set, which allows us to see whether IQ passes Leamer's extreme bounds test: If the value is positive, it passes.

Finally, in order to provide a more transparent measure of the extent to which IQ is robust and has the expected (positive) relationship with conditional growth, we also report the percent of regressions where β_{IQ} was positive and statistically significant at the 95% level as well as the percentage of regressions where the coefficient is simply positive, regardless of significance level. We do so in order to provide useful information about IQ's robustness to readers who are not sympathetic to Bayesian methods.

MATLAB software and data are available upon request, and online at www.siue.edu/∼garjone.

5. Estimation results

5.1. Robust regressors

Table 3 reports our main results. The first two rows report information on the IQ coefficient using data from all countries, while the third and fourth rows repeat these regressions, while omitting observations from the OECD countries. In the text, we focus attention on the BACE results and the percent significant results; the unweighted results tell much the same story.

Consider the first row of results: Using data from all countries, and including all possible 3-variable combinations of Sala-i-Martin's top 21 growth variables as explanatory variables (along with log GDP per capita in 1960, primary school enrollment in 1960, average lifespan in 1960, and a constant), IQ has a posterior inclusion probability of 96.1%, which reflects the fact that models that included IQ performed far better (in a BACE-weighted SSE sense)

Table 3 IQ's relationship with economic growth, 1960–1992

	$\overline{\beta}_{\text{IQ}}$ (unweighted)	$\overline{\beta}_{\text{IQ}}$ (BACE)	$\beta_{\text{IQ}} - 1.96\sigma_{\text{IQ}}$ Lower Bound	% sig.	% pos.	# of regr.	Posterior incl. prob.
All countries Controls = Top 21	0.123 (0.025)	0.111 (0.025)	−0.021	99.8%	100%	1330	96.1%*
All countries Controls = Educ	0.150 (0.023)	0.153 (0.023)	0.092	100%	100%	56	100%*
Non-OECD Controls = Top 21	0.103 (0.048)	0.082 (0.046)	−0.166	62.8%	99.9%	1330	29.7%*
Non-OECD Controls: Educ	0.131 (0.043)	0.163 (0.043)	−0.006	96.4%	100%	56	90.4%*

Note: $\overline{\beta}_{\text{IQ}}$ represents the average across all regressions of the effect of a one-point increase in a nation's average IQ on average annual economic growth, in percent. Standard errors (unweighted and weighted averages across all regressions) are in parentheses. "Lower Bound" is the minimum value of lower bound of the 95% confidence interval across all regressions. "Percent significant" is the percent of regressions where IQ was statistically significant at the 95% level. Posterior inclusion probability is fully defined in the text; an asterisk indicates that the posterior inclusion probability is greater than the prior inclusion probability. In all regressions, log GDP per capita in 1960, primary school enrollment in 1960, and average lifespan in 1960 are included as additional explanatory variables. Tob 21 and Education control variables are listed in Table 1

than IQ-excluding models. Of course, the posterior of 96.1% far exceeds the prior of 18.2%, so following SDM, we call IQ a "significant" growth variable.

For a rough comparison, consider posteriors from SDM, who first created and used the BACE approach. SDM's top three variables were the East Asian dummy (82.3%), primary schooling in 1960 (79.6%), and the relative price of investment goods (77.4%). By contrast, SDM's four lowest ranking variables had posterior inclusion probabilities of 1.5% each. Clearly, the BACE approach can generate wide dispersion of posteriors, a useful feature for separating econometric wheat from chaff. Of course, one should keep in mind that SDM's results are not directly comparable to ours: Since their model tested 67 variables, their baseline model implied a tougher prior inclusion probability of 10.7%. Nevertheless, the 96.1% posterior inclusion probability for IQ indicates that IQ-inclusive models have a large amount of explanatory power conditioned on this robust-variable dataset.

Further, note that IQ is statistically significant at the 5% level in 99.8% of the 1330 regressions. Thus, IQ failed to reach conventional statistical significance in just three of these 1330 regressions.[2] The coefficient is always positive.

We also note that IQ is statistically significant at the 1% level in 98.5% of all regressions—all but 20. Since each of the "top 21" are included in 190 = 20!/(18!2!) regressions, this means that *regardless* of the "top 21" variable in question, IQ is significant at the 1% level in the vast majority of regressions that include that particular control variable.

In addition, IQ is significant at the 0.1% level in 92% of the regressions, and at the .0001% level in 61.43% of the regressions. Thus, most of the time, IQ's statistical significance is far beyond the thresholds employed in most econometric research.

The third column reports the lowest value of the lower end of the confidence interval from all 1330 regressions; this is the value that must be strictly positive in order to pass Leamer's extreme bounds test. The value, −0.0214, means that IQ fails Leamer's rigorous test in this case (a result implied by IQ's statistical insignificance in three of the regressions).

[2] Two of the three were significant at the 10% level, and the third at the 18% level.

However, IQ passes the pseudo-t test quite easily. As noted above, SDM note that this will be valid as long as the posterior distributions of β_{IQ} "are not too far from being normal." As it happens, the weighted average IQ coefficient is more than four weighted average standard deviations from zero, so not only can we be extremely confident that the true coefficient is not zero, but—assuming we are not "too far from being normal"—we can also be 95% confident that the true value lies between 0.16 and 0.062. Thus, interpreting the model structurally, raising a nation's IQ by 10 points is estimated to add between 0.62% and 1.6% to a nation's annual growth rate of GDP per capita, with a point estimate of 1.11%. Below, we consider a more plausible interpretation of these growth rates in light of economic theory, seeing them as likely reflecting higher steady state productivity levels.

A histogram of the estimated IQ coefficients from this set of results is provided in Fig. 4; note that all estimated coefficients are positive, over 95% of them are greater than 0.09, and the distribution does not appear "too far from being normal." The median estimate is 0.124, which implies that raising a nation's IQ by 10 points is estimated to add 1.24% to the nation's annual growth rate.

We now move to the third row of Table 3, which omits data from all OECD countries. We view these results merely as suggestive, since in excluding OECD countries, we are thereby excluding roughly half our sample. In the non-OECD sample, the BACE-weighted standard error is less than two standard deviations away from the mean—the pseudo-t-statistic is 1.78, short of the mark. If we were, with appropriate caution, to interpret this as a conventional t-statistic, it would imply statistical significance at the 10% level.

These weaker results are likely due to restriction of range, since if the strong IQ results were in fact driven by the inclusion of the OECD countries, then we would not expect the following non-OECD results: The posterior inclusion probability is substantially greater than

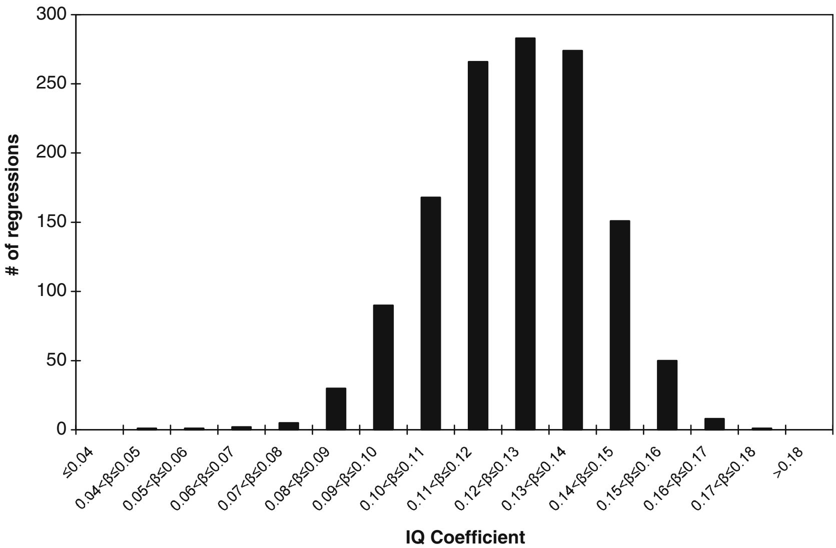

Fig. 4 Relative frequency of β_{IQ}
Note: Reflects relative frequency of β_{IQ} from 1330 regressions that included all possible combinations of the 21 growth variables that passed Sala-i-Martin's (1997) robustness test (see Table 1). Data from all available countries were included. In addition, log GDP per capita in 1960, primary school enrollment in 1960, and average lifespan in 1960 are included as additional explanatory variables in all regressions

18%, IQ is statistically significant in over three-fifths of the regressions, and IQ is positive 99.9% of the time.

Finally, a word about changing the model priors. We focus this discussion on the first row of Table 3, which includes data from all countries in the dataset. Note that if we sum (1) over all IQ-inclusive models, the result is the posterior inclusion probability for IQ-inclusive models, which we denote $P(\mathrm{IQ}|y)$. If in the denominator we separate IQ-inclusive models from IQ-excluding models this yields:

$$P(\mathrm{IQ}|y) = \sum_{i=1}^{1330} P(M_i|y)$$
$$= \frac{\sum_{i=1}^{1330} P(M_{\mathrm{IQ}})T^{-k/2}SSE_i^{-T/k}}{\sum_{m=1}^{1330} P(M_{\mathrm{IQ}})T^{-k/2}SSE_m^{-T/k} + \sum_{n=1}^{5985} P(M_{\sim\mathrm{IQ}})T^{-k/2}SSE_n^{-T/k}}.$$

We have already derived values for $P(\mathrm{IQ}|y)$ conditioned on $P(M)$ in Table 1. Before, we forced the prior for each model to be equal, whether or not IQ was included in the model, implying that $P(M) = P(M_{\mathrm{IQ}}) = P(M_{\sim\mathrm{IQ}}) = 1/7315$.

Now we relax that assumption, choosing any prior we like for $P(M_{\mathrm{IQ}})$; calculating the appropriate conforming value for $P(M_{\sim\mathrm{IQ}})$, models that exclude IQ, is an exercise left to the reader. Relaxing this assumption of equal priors seems especially appropriate since national average IQ is a relatively new growth variable. It is then straightforward to see how $P(\mathrm{IQ}|y)$ is impacted by changes in $P(M_{\mathrm{IQ}})$. In particular, we can drive our IQ prior, $P(M_{\mathrm{IQ}})$, as low as we like and observe how that impacts $P(\mathrm{IQ}|y)$, our Bayesian posterior inclusion probability for IQ-inclusive models.

We shall report these results in the variable inclusion framework used earlier. Recall that initially, the prior inclusion probability for IQ was 18.2%. If we drop this value to 10%, $P(\mathrm{IQ}|y) = 92.5\%$; if the IQ prior drops to 1%, the IQ posterior falls to 53%. So even if we had a prior belief that there is only a one percent chance that IQ belongs in the model, we would revise our beliefs after looking at the data, concluding that we should probably incorporate IQ into future growth models.

One question of particular interest is how low $P(M_{\mathrm{IQ}})$ could go before IQ would become a "below average" variable. Since the "average" variable will be present in 18.2% of all 7315 regressions, then what we would want to know is how low P(M) would have to be for $P(M_{\mathrm{IQ}}|y)$ to fall below 18.2%. The answer is 0.2%. Therefore, under the specification presented here, if one's prior for IQ-inclusive models is above 0.2%, then observing the results presented here should lead one to conclude that IQ is a significant growth variable.

Overall, we interpret the results from this section as providing strong confirmation of IQ's robustness. In light of the many criticisms of IQ tests, one might have expected that the IQ-growth link would be merely epiphenomenal, something that would fade away when researchers controlled for variables that were strongly related to growth. We did not find this to be the case.

5.2. Education regressors

Table 3 indicates that IQ easily passes a Leamer-style extreme bounds test when the particular set of education measures included here (listed in Table 1) are used as control variables. Out of our 56 education regressions using data from all 51 countries, the extreme lower bound across all regressions was still positive. Thus, the support for β_{IQ} appears to be strictly positive when other education variables are included as explanatory variables in the full-country dataset. Of

Table 4 Primary School Enrollment's Relationship with Economic Growth, 1960–1992

	$\overline{\beta}_{PS\%}$ (unweighted)	$\overline{\beta}_{PS\%}$ (BACE)	$\beta_{PS\%} - 1.96\sigma_{PS\%}$ Lower bound	% sig.	% pos.	No. of regr.	Posterior incl. prob.
All countries Controls = Top 21 IQ Included	1.52 (1.26)	1.93 (1.09)	−3.55	7.5%	98.0%	1330	21.8%*
All Countries Controls = Top 21 IQ Excluded	2.04 (1.60)	0.69 (1.39)	−4.62	11.5%	97.9%	1330	6.3%

Note: $\overline{\beta}_{PS\%}$ represents the average across all regressions of the effect on growth, in percent, of moving from 0% to 100% enrollment of the primary-school-aged population. Standard errors (unweighted and BACE-weighted averages across all regressions) are in parentheses. "Percent significant" is the percent of regressions where primary schooling was statistically significant at the 95% level. Posterior inclusion probability is fully defined in the text; an asterisk indicates that the posterior inclusion probability is greater than the prior inclusion probability. In all regressions, log GDP per capita in 1960 and average lifespan in 1960 are included as additional explanatory variables

course, the SDM pseudo-t statistic for IQ is quite large—6.6—and the inclusion probability is 100%. Therefore, in this dataset, no three-variable combination of education measures can eliminate the statistical robustness of IQ.

As before, we rerun these education results while omitting all OECD countries. As before, we view these results merely as suggestive, since in excluding OECD countries, we are thereby excluding roughly half our sample. But when we do exclude OECD countries, we find that the IQ coefficient is always positive and is statistically significant at the 5% level in 96.4% of all regressions—and at the 1% level in 62% of the regressions. The mean coefficient size is little affected when OECD countries are excluded.

These results are surprising considering the enormous weight placed in the development literature on quantitative education measures—one might thereby have reasonably expected national average IQ to be largely (or entirely) a proxy for some appropriately weighted education index. But we did not find evidence for that hypothesis, at least with the data and parameterizations considered here.

One question that naturally arises is how the inclusion of IQ impacts the estimates for Sala-i-Martin's best-performing human capital measure: primary school enrollment. Table 4 reports the results of such an inquiry. We only report results that include all 1330 combinations of Sala-i-Martin's 21 robust growth regressors. In Sala-i-Martin's original paper, primary school enrollment was statistically significant in 47.6% of the 32,509 regressions ($59!/(3!56!) = 32,509$). The second row, third column of the results from Table 4 shows that even when IQ is excluded from our regressions, primary school enrollment is statistically significant in only 11.5% of the 1330 regressions, and the posterior inclusion probability is well below the 18.2% prior. This is most likely due to the fact that only robust regressors are included in the regression: when weak regressors are excluded, the explanatory power of primary school enrollment diminishes (We should note that the reduced significance of primary schooling could also be due to the smaller sample size in our database). In no case is the primary school enrollment coefficient twice the size of the BACE standard error, and in the IQ-excluding case it is smaller than the BACE standard error.

But the evidence is not unambiguously against the inclusion of primary schooling. When IQ is included in the regressions, primary school enrollment clears the prior inclusion probability threshold, and reaches a posterior inclusion probability of 21.8%. Further, the primary

Table 5 Growth variables passing the robustness test of SDM (2004), in rank order

1. East Asian dummy +
2. Primary schooling 1960 +
3. Investment price −
4. GDP 1960 (log) −
5. Fraction of tropical area −
6. Population density coastal 1960's +
7. Malaria prevalence in 1960's −
8. Life expectancy in 1960 +
9. Fraction Confucian +
10. African dummy −
11. Latin American dummy −
12. Fraction GDP in mining +
13. Spanish colony −
14. Years open to trade +
15. Fraction Muslim +
16. Fraction Buddhist +
17. Ethnolinguistic fractionalization −
18. Government consumption share 1960's −

Note: + and − signs indicate whether more of that value appears to be good or bad for economic growth, according to Table 2 of SDM (2004).

school enrollment coefficients are positive roughly 98% of the time, so while primary schooling's signal-to-noise ratio may be high, there is at least clear evidence regarding which direction the signal is pointing.

In results not reported here, we estimated sets of regressions similar to those reported in Table 4, but used *all* education measures (one at a time) as the schooling variable. Our intention was to assess the marginal importance of IQ in the growth regressions. However, as in Sala-i-Martin's work, primary school enrollment was the most robust variable among all of the education-related measures we tested. The other education measures were so weakly correlated with growth that we do not report the results here.

This dramatic decline in the statistical significance of primary school enrollment once weak regressors are excluded makes the performance of IQ—statistically significant in 99.8% of the same regressions—all the more surprising. Not only is IQ robustly correlated with economic growth in this sample: it is also the most robust human capital measure in this dataset.

5.3. Robustness check: Data from SDM (2004)

As an additional robustness check, we reemployed the methodology from Section 5.1 with a new set of robust regressors: The 18 growth variables that passed SDM's BACE test. These variables are listed in Table 5. The dependent variable is the growth rate of GDP per capita between 1960 and 2000. One strength of this dataset, which has 88 observations, is that SDM searched for the combination of countries and variables that minimized the number of missing observations. Our IQ data overlaps with a maximum of 54 of the observations.

As before, we include seven control variables (plus a constant) in each regression. The first is, of course, national average IQ. The three fixed variables are the three specifically noted by SDM as having the "strongest evidence. . .the relative price of investment, primary school enrollment, and the initial level of real GDP per capita" (p. 813). We include these in all regressions since the evidence in their favor is so strong that omitting them from any regression would likely imply model misspecification.[3] We then include all three-variable combinations of the remaining 15 robust regressors, for a total of $15!/(12!3!) = 455$ regressions.

The results can be stated quite simply: IQ passes Leamer's extreme bounds test at the 5% level as well as at the 1% level. In other words, IQ is always statistically significant at conventional levels when conditioned on these robust regressors in these 455 regressions. The unweighted mean coefficient is 0.115, and the unweighted mean standard error is 0.022. 98.7% of the regressions are significant at the 0.1% level, and 64% at the 0.001% level. These results are sufficiently robust that we will refrain from reporting BACE-weighted summary statistics and inclusion probabilities.

6. Discussion

Overall, our results indicate that higher IQ is associated with an economically large and statistically significant increase in growth rates. It is reasonable to wonder what these results mean. One might interpret these results as indicating that IQ measures a key output of the education, socialization, and child-rearing process—an output called general reasoning ability—while primary school enrollments are a measure of one key input into this human-capital-creation process. Inputs are likely to have a noisy relationship with outputs, so the weak relationship between schooling and economic performance is little surprise.

What is a surprise, at least from the point of view of much growth research, is that a heretofore overlooked measure of human capital—the IQ test—is so robustly related to growth. Growth economists may know little about how a nation's stock of human capital is produced,[4] but it appears that we at least have a tool for measuring a critical portion of that stock of human capital.

Further research can now be done to determine exactly what role this form of human capital plays in the growth process. Is national average IQ an engine of growth, part of the technology production function from a Romer-type endogenous growth model? Or perhaps a high national average IQ is more critical as a resource for *adapting* the technologies developed elsewhere, a role played by human capital in Bils and Klenow (2000) and Jones (2002, c.6).

More broadly, does IQ have an affect on the *growth rate* of living standards in steady state—something possible in endogenous-growth models—or is it more likely to have an effect only on the *level* of steady state living standards—as in a Solow-style model with human capital accumulation? Fully developing and testing models that could answer this question is far beyond the scope of this paper, but the results here should provide an impetus to future work.[5]

But to give a sense of the quantitative importance of cross-country IQ differences, let us take a moment to interpret our regression results within the standard Solow/Mankiw—

[3] Table 5 notes that the East Asian dummy is SDM's best-performing variable in their baseline specification, but SDM note that the three variables mentioned above perform slightly better when a wider variety of specifications are considered.

[4] Galor and Moav (2002) provide an intriguing Darwinian explanation for parental investment in children's human capital accumulation. In their model, the Malthusian era favors the transmission of pro-human-capital-investment genes. They focus on the experience of Western Europe; extending their model to the rest of the world could help explain why average IQ differs so much across countries.

[5] Jones (2006) shows that in a calibrated Ramsey model, IQ differences alone can explain up to half of the world income distribution.

Romer–Weil framework. Here, IQ impacts the steady-state growth path—informally, the "level" of living standards—but has no permanent effect on the steady-state growth rate. It is straightforward to reinterpret results from our growth regressions in this light. Jones (2000) and Barro and Sala-i-Martin (2001, p. 466ff.) demonstrate how to do so. Consider the following representation of the economy of country *i*:

$$\Delta \log y_{it} = g + \beta(\boldsymbol{\gamma}'\mathbf{A_i} - \log y_{i0}) + \varepsilon_{it} \tag{2}$$

where, as usual, y_{it} and y_{i0} indicate per capita GDP during the current period and at the beginning of the period and g is the exogenous long-term growth rate. The term in brackets reflects the gap between the steady-state log GDP level and the starting point of log GDP; the Solow model implies that when the gap is bigger, growth will be faster.

The first term within brackets deserves further attention. The (k $\times$ 1) vector $\mathbf{A}_i$ contains the country-specific levels of various institutional, cultural, geographic, and other variables that would have an impact on the steady-state log-level of GDP. The (1 $\times$ k) vector $\boldsymbol{\gamma}'$ is then the vector of parameters that summarize how these variables impact the steady-state level of GDP per capita. Therefore, the product of these two vectors is then the steady-state level of log GDP per capita.

Therefore, if (2) represents the true data-generating process, then when economists run growth regressions, the variables we typically call "growth regressors" are really not providing information about the steady-state growth rate, since growth is exogenous to the model and shows up in the constant. Instead, the growth regression coefficients we estimate are really the (1$\times$k) vector $\beta\boldsymbol{\gamma}'$.

Accordingly, it is the vector $\boldsymbol{\gamma}'$ that contains all information about how changes in the economy's parameters (its education level, savings rate, latitude, etc.) impact the *level* of steady-state living standards. This information can be easily recovered by dividing $\beta\boldsymbol{\gamma}'$ by β, the coefficient on the starting level of log GDP per capita. The element of $\beta\boldsymbol{\gamma}'$ that matters for our purposes is the IQ coefficient. The first line of Table 3 indicates that the unweighted average IQ coefficient is 0.12; we found that the coefficient for starting log GDP per capita averaged 1.98.[6] This is extremely close to the 2–3% value that Barro and Sala-i-Martin (2001, p. 496, p. 521) take as canonical.

These values imply that γ_{IQ}, the effect on log GDP per capita of a rise in national average IQ of one point, is 0.061. In other words, assuming that (2) is the true data-generating process, one IQ point appears to raise steady-state living standards by 6.1%. This is about 2/3 of the value implied by the simple bivariate relationship between IQ and log GDP per capita illustrated in Table 1, where one IQ point was associated with an 8.7% rise in log GDP per capita. Thus, even after controlling for the most robust growth variables, the relationship between IQ and log GDP per capita is mostly intact.

Finally, for an overall assessment of how IQ compares to other common growth variables, consider Sala-i-Martin's original results, which used combinations of 62 growth variables in over two million regressions. Among his top 21 regressors—the ones which he considered robust—the median regressor was statistically significant in 76.4% of cases, with a range from 100% (for fraction Confucian) to 2.81% (for revolutions and coups). Fraction Confucian was the only regressor that passed an extreme bounds test at the 5% level. Only eight

[6] Note that since, in our BACE framework, both β and $\boldsymbol{\gamma}'$ are weighted the same way, our answer is invariant to using BACE-weighted versus unweighted coefficients.

of his top 21 had coefficients over three standard errors from zero, while in our full-sample results using his top 21 growth variables, IQ's coefficient is over four standard errors away from zero. IQ would thus appear to fit comfortably in the top half of Sala-i-Martin's top 21 growth variables.

7. Conclusion

If human capital is important in economic development, then it would be valuable to have a reliable measure of at least some portion of this stock of human capital. The evidence presented here indicates that general intelligence as measured by IQ tests is a reliable indicator of one important form of human capital. We further show that such general-purpose human capital has a statistically robust and economically large positive correlation with economic growth. IQ easily outperforms the best-performing measure of human capital in Sala-i-Martin's widely used dataset—primary school enrollment—and is statistically significant in all but three out of 1330 full-sample growth regressions. Even when OECD countries are excluded from the sample, IQ appears to have an economically large and statistically significant positive relationship with growth. And when IQ is included in 455 regressions that include all of the robust growth regressors from SDM (2004), IQ is statistically significant at the 1% level in all regressions.

Considering the many criticisms that IQ tests receive, their robustness in these growth regressions is all the more surprising. It would, of course, be extremely valuable to have IQ data from more countries over a longer time period. We hope that these encouraging results encourage the collection of cross-country IQ data in the future.

There is one critical issue we have mentioned here but have not fully addressed: The endogeneity of IQ over time. We mentioned the Flynn effect, the 2–3 points-per-decade increase in IQ found in developed countries, an increase that appears to come mostly from a rise in the bottom of a population's IQ distribution. This effect gives researchers some reason to believe that increases in the quality and quantity of education, reductions in poverty, and increases in overall literacy can increase a nation's average IQ.

Our estimates of IQ account for the Flynn effect, but do so imperfectly. In particular, psychologists are just beginning to understand why the Flynn effect is higher in some countries rather than others, so we do not make country-specific Flynn effect adjustments to our IQ data. But as the structure of the Flynn effect becomes clearer, economists and psychologists may uncover Solow-type convergence results for national average IQ.

We hope that the results presented here will encourage growth and development economists to join this area of research. The stakes appear to be quite large. For instance, if the relationship documented here proves to be causal, then one IQ point raises steady-state living standards by an estimated 6.1%. Since estimates of national average IQ across countries span a range of greater than 30 points, it appears that IQ differences alone can explain a substantial fraction of cross-country differences in living standards.

A key lesson here appears to be that the health and vigor of the human brain is likely to be a key determinant of national economic performance. Economists who wish to understand the large, persistent differences in cross-country economic performance seen in the data will need to devote at least some energy to understanding why IQ differences across countries are so very large. Economists and policymakers who wish to reduce cross-country gaps in living standards will need to devote energy to understanding what can be done to narrow those cross-country IQ differences; the psychological literature surveyed in Section 2 serves as an introduction to these issues. If economists bring their powerful theoretical and econometric tools to bear on the relationship between IQ and economic performance, one can only hope that these crucial questions will find promising answers.

Acknowledgements The authors would like to thank Richard Lynn, Tatu Vanhanen, Julie Hotchkiss, Garey Ramey, Joseph Haslag, Rik Hafer, Ping Wang, two anonymous referees, the editor of this Journal, and participants at the Econometric Society's 2004 Latin American meetings, the 2004 Eastern Economic Association meetings, the Econometric Society's 2005 World Congress, and the University of Missouri's departmental seminar for many very helpful comments. Jones would also like to thank Teodora Borota for outstanding research assistance, and the Graduate School of Southern Illinois University Edwardsville for financial support. The usual disclaimer applies.

References

Armor, D.J. (2003). *Maximizing intelligence*. New Brunswick: Transaction Publishers.

Barro, R., & Sala-i-Martin, X. (2001). *Economic Growth* (2nd ed.). Cambridge, MA: MIT Press.

Barro, R.J., & Lee, J.-W. (1994). Sources of economic growth, *Carnegie Rochester Conference Series on Public Policy*. 40, June 1994.

Barro, R.J., & Lee, J.W. (1996). International measures of schooling years and schooling quality, *American Economic Review, 86*(2), 218–223.

Barro, R. J., & Lee, W. (1993). International comparisons of educational attainment, *Journal of Monetary Economics, 32*, 363–394. Data available at www.nber.org.

Bils, M., & Klenow, P. (2000). Does Schooling Cause Growth or the Other Way Around? *American Economic Review, 90*, 1160–1183.

Bleaney, M., & Nishiyama, A. (2002). Explaining Growth: A Contest between Models, *Journal of Economic Growth*, March, 7, 43–56.

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

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since *Bias in Mental Testing*, *School Psychology Quarterly, 14*, 208–238.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Court, J. H. (1991). Asian applications of Raven's Progressive Matrices, *Psychologia, 34*, 75–85.

Daley, T. C., Whaley, S. E., Sigman, M. D., Espinosa, M. P. & Neumann, C. (2003). IQ on the rise: The Flynn effect in rural Kenyan children, *Psychological Science, 14*, 215–219.

Flynn, J. R. (1987). Massive IQ gains in 14 nations, *Psychological Bulletin, 101*, 171–191.

Galor, O., & O. Moav (2002). Natural selection and the origin of economic growth, *Quarterly Journal of Economics, 117* (November), 1133–1191.

Gottfredson, L. (1997). Why g matters: The complexity of everyday life, *Intelligence, 24*, 79–132.

Gottfredson, L. (2004). Intelligence: Is it the epidemiologists' elusive 'fundamental cause' of social class inequalities in health? *Journal of Personality and Social Psychology, 86*, 174–199.

Gould, S. J. (1981). *The mismeasure of man*. New York: W. W. Norton/Harmondsworth, U.K.: Penguin Books.

Hanushek, E. & Kim, D. (1995). Schooling, labor force quality, and economic growth, National Bureau of Economic Research (Cambridge, MA), Working Paper No. 5399, December.

Hanushek, E. & Kimko, D. (2000). Schooling, labor force quality, and the growth of nations, *American Economic Review, 90*, 1184–1208.

Herrnstein, R. J. & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*, New York: Free Press.

Hoover, K. D., & Perez, S. J. (2004). Truth and robustness in cross-country growth regressions, *Oxford Bulletin of Economics & Statistics, 66*(5), 765–798, December.

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, CT: Praeger.

Jones, C. (2000). Comment on Rodıguez and Rodrik, 'Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence.' Manuscript, University of California.

Jones, C. (2002). *Introduction to economic growth.* New York, NY: W.W. Norton.

Jones, G. (2006) IQ in the Ramsey Model: A Naïve Calibration, Working paper, Southern Illinois University Edwardsville.

Koop, G., & S. Potter (2003). Forecasting in large macroeconomic panels using Bayesian Model averaging, *Federal Reserve Bank of New York Staff Reports* no. 163, March.

Leamer, E. E. (1978) *Specification searches*. New York: John Wiley & Sons.
Leamer, E. E. (1983). Let's take the con out of econometrics, *American Economic Review, 73*, 31–43.
Leamer, E. E. (1985). Sensitivity analysis would help, *American Economic Review, 75*, 31–43.
Levine, R., & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions, *American Economic Review, 82*, 942–963.
Lynn, R. (1987). The intelligence of the Mongoloids: A psychometric, evolutionary, and neurological theory, *Personality and Individual Differences, 8,* 813–844.
Lynn, R., & T. Vanhanen (2002). *IQ and the wealth of nations*. Westport, CT: Praeger Publishers.
Mankiw, N. G., Romer, D. & Weil, D. (1992). A contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics, 107*, 407–438.
Milani, F. (2004). Monetary policy with a wider information set: A Bayesian Model averaging approach, Working Paper, Princeton University.
Murray, C. A. (1997). IQ and economic success, *The Public Interest, 128*, 21–35.
Murray, C. A. (2002). IQ and income inequality in a sample of sibling pairs from advantaged family backgrounds, *American Economic Review*, *92*, 339–343.
Neal, Derek A., & William R. Johnson. (1996). The role of premarket factors in black–white wage differences, *Journal of Political Economy, 104*(5), 869–895.
Neisser, U., et al. (1996). Intelligence: Knowns and unknowns, *American Psychologist, 51*, 77–101.
Ree, M. J., Carretta, T. R., & Green, M. T. (2003). The ubiquitous role of g in training, In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 261–274). Amsterdam: Pergamon.
Reed, T. E. & Jensen, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence, 16*, 259–272.
Rushton, R. J., & Rushton, E. W. (2003). Brain size, IQ, and racial-group differences: Evidence from musculoskeletal traits, *Intelligence, 31*, 139–155.
Sailer, S. (2004). IQ and the wealth of nations, Lynn and Vanhanen: Data table of national mean IQ studies. http://www.isteve.com/IQ_Table.htm
Sala-i-Martin, X. (1997a). I just ran two million regressions, *American Economic Review*, *87*(2), 178–183. Data available at www.columbia.edu/∼xs23.
Sala-i-Martin, X. (1997b). I just ran four million regressions, National Bureau of Economic Research Working Paper 6252.
Sala-i-Martin, X. Doppelhofer, G. & Miller, R. (2004). Determinants of long-run growth: A Bayesian averaging of classical estimates (BACE) approach, *American Economic* Review, *94*(4), 813–835.
Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, March, *6*(2), 461–464.
Seligman, D. (1992). *A question of intelligence: The IQ debate in America*. New York: Birch Lane Press.
Sternberg, R. J., & Grigorenko, E. L. (2001) *Enviromental effects on cognitive abilities.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
Temple, J. (2000). Growth regressions and what the textbooks don't tell you, *Bulletin of Economic Research, 52*, 181–205.
Volken, T. (2003). IQ and the wealth of nations. A critique of Richard Lynn and Tatu Vanhanen's recent book, *European Sociological Review, 19*, 411–412.
Weede, E., & Kampf, S. (2002). The impact of intelligence and institutional improvements on economic growth, *Kyklos, 55*, 361–380.
Winship, C., & Korenman, S. (1997). Does staying in school make you smarter? The effect of education on IQ, In *The Bell Curve*, In *Intelligence, Genes, and Success: Scientists Respond to the Bell Curve.* New York, NY: Springer-Verlag.
Wright, J. (2003). Forecasting U.S. inflation by Bayesian Model Averaging. *International Finance Discussion Papers,* 780, Board of Governors of the Federal Reserve System.
Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: John Wiley & Sons.