The Journal of Economics and Politics

Volume 27 | Issue 1 Article 2

2022

The Methodology of Multiple Regressions for GDP per Capita for Nations in the World Between 0 CE and 1940

Robert Rogers
Ashland University, rrogers1@ashland.edu

Recommended Citation
Rogers, Robert (2022) "The Methodology of Multiple Regressions for GDP per Capita for Nations in the World Between 0 CE and 1940," *The Journal of Economics and Politics*: Vol. 27: Iss. 1, Article 2.
Available at: https://collected.jcu.edu/jep/vol27/iss1/2

The development of the Maddison datasets of world-wide Gross Domestic Product, GDP**,** has given scholars the opportunity to analyze national income models over long time periods (0CE to the present, for instance). Besides the Maddison GDP data, a number of other variables such as population, climate attributes, political characteristics, and location are available for long time periods.

I develop and estimate here a series of models of GDP per capita (GDPC**)** for a sample of countries in Eurasia, Africa, Oceania, and the Americas for the available years between 1 BCE and 1940. How to go about this task requires some thought and scrutiny. Thus, this paper focuses on the methodological approach rather than the precise determination of the results. The standard method of estimating this type of model develops a reduced form equation relating growth as depicted by GDPC to a set of exogenous variables. Tests are, then, made (usually t tests) to see if given variables have a sign and size consistent with theory. Variables without the significant signs are usually abandoned, and the variables with the statistically significant signs are hypothesized to be the most plausible parts of the model. While variables that are thought to be compellingly important are sometimes included even if they are not statistically significant, the models chosen usually contain only statistically significant variables. This method has been called the classical approach.

An alternative methodology is the Bayesian approach whereby a large number of models are estimated, and the parameters are examined to see whether given variables exert the expected influence for a wide range of model specifications. The two most used Bayesian methods are Extreme Bounds Analysis (EBA) and Bayesian Analysis of Classical Estimation (BACE). The first method ascertains whether given variables are statistically significant in the predicted direction under a wide range of specifications, and the second estimates a weighted average of the parameters for the total range of models to see if they are on average significantly different from zero in the expected direction.

For this paper, I estimate a plausible GDPC model by first the classical method and then by the two Bayesian methods. After that, I compare the results. In this way, one may be able to assess the adequacy of the classical method.

I do not focus on the Full Structural Model of GPDC determination; rather I estimate the reduced form model which directly connects the important available independent variables in the model to GDPC. This approach is the best way to estimate the final direction and impact of given exogenous variables on GDPC.

Below, I first discuss the literature behind the model. Second, I set up the empirical model, deriving the reduced form model from a plausible theoretical base. Third, I present, develop, and discuss the variables used in the model. Then, I present the empirical results, and last, conclusions are drawn.

Literature Review

In the 1990s, world events such as the fall the Soviet Union and the apparent triumph of market institutions in other countries led to an increased interest in economic growth. A major problem encountered is the possible constraint that the physical capacity of an economy can put on GDPC. Some scholars have asserted that the coming of the scientific and industrial revolutions lessened (and maybe even eliminated) the physical constraints on the economy (Mokyr, 2017). In this paper, I apply the model used in this literature to the economies existing in the period before the industrial revolution and some years after, up to 1940. It was a period when most countries had experienced some change. This paper’s goal is to find ways to ascertain variables that had a significant impact.

The 1990s developments in economic growth theory have three characteristics. First, there was a renewed effort to apply the aggregate production function first developed by Solow (1957 and 1994). One of the major protagonists of this movement was Barro, and a textbook written by him and Sala-i-Martin, (Barro & Sala-i-Martin, 2004) has had a great influence. The literature discusses the change from an economy based on the Malthusian limitations to a more unconstrained society.

The second characteristic of this literature was the recognition that the orthodox econometric techniques may not have been appropriate, due to the lack of clarity about the underlying models. Thus, using the orthodox econometric techniques could lead to misleading results. This recognition has led to the application of Bayesian methods to the model and data. Here, however, I want to step back a little and make a comparison between using the orthodox approach and applying Bayesian techniques.

The third characteristic of this growth interest was the recognition that factors other than the usual strictly economic variables such as physical capital and technology could affect economic growth rates and national Gross Domestic Product (GDP). This paper includes variables on three of these causes; geographic location, political and social factors, and population genetics. Consequently, I examine some of the literature that suggests ways to estimate such a model.

A number of papers develop theories and economic models that could be applied to countries in the period prior to the industrial revolution. Van Zanden and van Leeuwan (2012) examine the history of GDP in Holland or the Netherlands in the period between 1347 and 1874. They find that economic growth was occurring during this period though it slowed down after 1650. Erdkamp (2016) examines the economic performance of the Roman Empire. He provides evidence for the proposition that political developments not Malthusian population limits led to economic decline in the area once ruled by Rome. Galor coauthored a series of papers (Galor & Moav, 2006; Galor & Weil, 2000; and Galor, et al., 2009) showing that social development and technological progress could overcome the physical limitation posed by Malthus (1798/1976).

Scholars have used various methods to examine the addition of many political and social variables to the Solow (1957 and 1997) production function to determine the causes of economic growth. (McGuire & Olson, 1996; Owen et al., 2009; Johnson & Koyama, 2017; Moreno & Trehan, 1997; Gassebner et al., 2016; and Auer, 2013.)

Given the nature of the situation, there can be a lack of rigorous theorizing behind the specification of the econometric models depicting national income growth. Often specifying these models leads researchers to set up very specific theories. Thus, the empirical analysis sometimes depends on assumptions that cannot be verified, and consequently, errors in the assumptions can lead to biased estimates. These erroneous preconceptions present an opening for the Bayesian methods.

An interesting set of papers uses Bayesian methods to ascertain the variables which had the most impact on the economic growth. These methods limit the number of maintained assumptions about the structure of the model: thereby increasing the robustness of the findings to differences in this structure. Two of the techniques used are Extreme Bounds Analysis (EBA) and Bayesian Averaging of Classical Estimates (BACE). One of the first of these papers is Levine and Zervos (1993); they apply EBA to samples from earlier works. Levine and Renelt (1992) using EBA finds that investment and trade factors are robust, but certain political variables are not. Sturm and Haan (2005) use various robust methods to model economic growth between 1960 and 1990. They stress that different estimation techniques can lead to result different from the conventional methods.

The papers using the Bayesian method that have had the most influence on this present work are Sala-i-Martin (1997) and Sala-i-Martin, Doppelhofer, and Miller (2004). (See also chapter 12 in Barro & Sala-i-Martin, 2004.) BACE estimates models for all conceivable combinations of the independent variables being considered. Through averaging and examining the results from different configurations, a researcher can determine which variables have the most empirical support in a large number of models. This effort provides answers to the questions as to which results are the most robust.

This process entails estimating a large number of models. For instance, if there are k variables under scrutiny, one must estimate the 2^k models to examine their impact under all circumstances. In my analysis, there are 10 BACE variables implying 1024 equations. Often, researchers use so many variables that they can only examine a sub-set of available models. For instance, Sala-i-Martin et al. (2004) had 67 variables that implied 2^{67} equations. (As an exercise, the reader may want to see how large this number is.) Sala-i-Martin et al. (2004) based their analysis and conclusions on a random sample of the possible BACE models. In this paper, however, I use the universe of possible explanations from the proposed variables because estimating them all is feasible.

Using a late 20th century sample, the Sala-i-Martin et al. (2004) found a total of eighteen variables to be robust in that they were significant and sizable over a large number of specifications. Among those are the conventional economic and sociological and political variables: examples being investment price, 1960 GDP (pre-sample), an East Asia dummy variable, and the population fraction that is Muslim. As discussed below, many if not most of the variables used in the above papers were either not available or irrelevant for the time periods in this paper's sample. This paper is a smaller version of the Sala-i-Martin et al. (2004) that it tests for 10 variables instead of 67. Furthermore, it applies both EBA and BACE analysis to proposed variables.

The Model - Structural and Reduced-Form Elements

As indicated above, it is reasonable to empirically estimate the connections between GDPC and the available causal variables even if not all the theoretical interconnections are known. To see the implications of this situation, I examine how a reduced form model such as I estimate can be derived from a theoretical structural system.

To illustrate, I examine this simplified structural model depicting the causation of two dependent variables, Y_1 and Y_2, which are determined by the independent variable vectors, X and Z:

$$Y_1 = f(Y_2, X) \qquad (1a)$$

$$Y_2 = f(Y_1, Z). \qquad (1b)$$

To explain the situation, I use two endogenous variables here even though I intend to estimate only one of the models. Often the information necessary to estimate this full model (1a and 1b) is not available, but it still may be possible to determine the impacts of given elements of X and Z on Y_1, and Y_2 using the reduced form. To get the reduced form, Y_1, and Y_2 are substituted into equations 1a and 1b arriving at:

$$Y_1 = r_1(X, Z) \tag{2a}$$

$$Y_2 = r_2(X, Z). \tag{2b}$$

With these equations, one can first estimate equations 2a or 2b and, second, evaluate the impacts of exogenous variables on the dependent variables, Y_1 and Y_2. The focus of the model is on only one of the endogenous variables. Here I used only Y_1. Consequently, only the reduced-form model of Y_1, which in this case is GDPC is estimated and analyzed.

Historically, this procedure has been the standard way to estimate the impact of one variable on another. With selected GDPC data from almost two millennia, however, there are special data problems. For many variables we do not have good proxies, and the exact value of the models is not clear.

To deal with this situation, a Bayesian analysis is deployed. It starts with the reduced form model above, equation 2a, which is the one being examined. No model assumptions are made as to which exogenous variables are included in the model. Essentially the model assumption is that it is possible that any combination of the exogenous variables can explain the behavior of Y_1. Thus, the model is estimated for every combination of the X and Z vectors. Two Bayesian methodologies are applied: Extreme Bounds Analysis (EBA) and Bayesian Analysis of Classical Estimator (BACE). With EBA, the coefficients of given variables are examined to determine whether they all have predicted sign for each given variable. That the coefficients for all the equation models have the predicted signs is evidence for the hypothesis that the variable's influence is in the predicted direction. The results for the sign are, then, said to be robust.

With Bayesian Analysis of Classical Estimator (BACE), weighted averages are computed for the coefficients of the relevant variables -- the weights being related to the probability value of the estimating equations. The t values are, then, computed for averages of the coefficients by dividing the weighted coefficient values by the weighted standard deviations. With these t values, one can determine if these coefficients are significantly different from zero in the direction predicted by the given theoretical hypothesis. The results from this exercise give us an indicator of the impacts of the independent variables in a wide range of situations. It, thus, shows a reasonable estimate of the impact of given variables under a wide range of possible models.

Reduced Form Variables in the Model

Having derived the reduced form model for GDPC for the period, 1BCE to 1940, the next step is to ascertain and define the variables in the model. In terms of the equation system derived in section III, the estimation model equals equation 2a. Essentially X and Z consist of all the available exogenous variables that impinge on GDPC, which is the equivalent to Y_1. Here, I now describe and discuss *these* X and Z variables.

The thesis behind this paper and the above-described model is that certain characteristics of a country can impact on per capita GDP in many different ways. To see the impact of these variables, I use the linear model where the dependent and all the independent variables are the raw numbers. With this model, the regression coefficient is the raw impact of both the continuous and dichotomous independent variables. The dependent variable is GDPC, the Gross Domestic Product per capita in millions of 2011 United States dollars from the Maddison data set (Bolt & Van Zanden, 2014). For each observation nation, this statistic has been adjusted for purchasing power parity. The geographic extent of the country observations for different observation years are developed and described by Bolt and Van Zanden.

The data or GDPC are available for only a sub-set of the sample. The included years are as follows: 0CE, 1000, 1500, 1600, 1700, 1820, 1850, 1870, 1900, 1913, 1920, 1925, 1930, 1935, and 1940. Unfortunately, the incompleteness of the dataset precludes testing some of the hypotheses that one might find interesting in the sample time period.

Now, I describe the variables that will be included in all renditions of the econometric model called above the X an Z vectors. The first representing the physical situation of a nation is its area, AREA. As discussed above, this variable has been discussed and defined by Bolt and Van Zanden (2014). More land area means more physical resources to produce GDP. Given the variation in this value over time, however, it is unlikely that simultaneity between physical size and GDPC is an issue.

Climate can also affect a country's productivity. LAT is the absolute value of latitude, which indicates the distance from the equator, and thus a good proxy for climate. Another variable representing the physical environment, ELEV, is the elevation of the central point of the country. This depicts the capacity of a country to produce output at given latitude with a given level of technology. In the far north or south higher elevation could lower productivity, while for countries near the equator it might increase productivity.

Since the most efficient way to move large quantities of goods is by sea, landlocked countries have less access to trade. This was especially true before the industrial revolution when the present day large mechanized ships were developed. Therefore, a dummy variable, LANDL, equaling one for landlocked countries and zero otherwise is included.

Because the sample consists of a panel data set, dummy variables for the included years, 0, 1000, 1500, 1600, 1700, 1820, 1850, 1870, 1900, 1913, 1920, 1925, 1930, and 1935 are included in the model. The base year is 1940. To account for many omitted variables, these dummies are included in all renditions of any given model.

I now describe the variables that have been suggested by the various hypotheses about the political, sociological, and genetic causes of economic growth in the world economy. Hypotheses on these variables are tested below by the use of not only the classical testing method but also the above-described EBA and BACE techniques. Recent work has shown that having colonies can increase a nation's GDPC; the ability of a nation to exploit its colonies apparently overcame the costs of colonization (Rogers, 2019). Therefore, I add to the model the variable, COL1, equaling one for countries with overseas colonies and zero, otherwise.

Additionally, the hypothesis that being a colony or part of a larger empire controlled by foreigners could change income is tested. This reflects that the possibility that the colonist country could be exploiting the colony. A dummy variable, COLO, is included in the model; it equals one for countries that were controlled by foreign nations and zero, otherwise.

Malthus (1798/1976) argues that the physical characteristics of a nation put a limit on its potential and actual national output. With modern technology, this limit has been greatly attenuated. In fact, it may be so attenuated that it has no or at least an insignificant impact. To depict this condition, I add to the model the variable, MAREA, equaling the product of AREA and one for the period before the industrial revolution and zero, otherwise. If this variable impact is negative and significant, then, support is given to the Malthusian hypothesis that before the industrial revolution the physical environment of a country put a greater limit on its GDPC than it had after the revolution when these limits were attenuated.

The theory called Efflorescence posits that some societies including ones before, during and after the industrial revolution have a greater ability to increase GDPC with improving technology. Societies with Efflorescence have or had a greater propensity to economic growth than others. In addition to the national observations after 1850 for most European and American nations, Australia, New Zealand, and Japan, the following earlier societies are hypothesized to have been Efflorescent: 11th and 12th century Sung Dynasty China, the Netherlands in the 17th and 18th centuries, 18th century Qing Dynasty China, Britain in the 1700s, and the area controlled by the Abbasid Caliphate in the year, 1000. Thus, I add the variable, EFFL, equaling one for all the countries with hypothesized to have Efflorescence and zero, otherwise. I also analyze a second variable, EFFL1, equaling one for the periods when certain countries are thought to have had Efflorescence before the industrial revolution, and zero, otherwise (Jones, 2007; and Goldstone, 2002).

A third theory has been suggested by the literature (Jones, 2007; and Karayalcin, 2008). It proposes a variable based on the bargaining relationship between each country's government and its subjects or citizens. The rulers provide the services from government in exchange for taxes. Accordingly, the subjects pay taxes for the desired (and/or maybe not desired) services provided by the government rulers. In most all situations, the government is a monopoly provider giving it considerable if not necessarily complete market power.

The control and power over subjects or citizens exerted by the government is constrained by a number of factors. First, some states have constitutions that limit their power. These states have less control than others. Second state managers often recognize that a too high price for state services can result in less revenue; thus, some situations call for restraint on taxes. But often rulers may not recognize these situations, and therefore, the characteristics of the ruler and the state can determine the relative bargaining position between the state and its subjects. (Sng, 2014.)

For many, one of the most relevant restraints on the state power is the ability of subjects to migrate to other countries or jurisdictions. If certain subjects find the government impositions too onerous, they will migrate. Often, these people are among the most economically efficient. Many scholars have posited that this has been a major factor in the economic success of Europe in the last millennium. The small size of European countries and the large number of countries in relatively small space mean that there are a large number of different governments in a relative short distance from each other. This situation gives the subjects opportunities to move, if they were not satisfied with the situation in their own country. This openness allows people to migrate more easily, if they find greater opportunities in other countries. This limits the ability of the government to constrain the behavior of its subject (Jones, 2007; Karayalcin, 2008; and Chaudry & Garner, 2006).

In this paper, I depict the situation with a variable representing the ability of the subjects and citizen to leave a given jurisdiction: NEIGHNO. This variable represents the subject's choice set; it equals the number of different independent governments within 1000 miles of the center of the observation country. The variable, thus, indicates the number of countries to which it is feasible for a subject to migrate if they are dissatisfied with their present situation. Such alternatives constrain the ability of a given government to tax subjects.

Therefore, the higher NEIGHNO the greater the immigration choice of subjects or citizens and the lower the cost of immigration. This variable by portraying the ability of subjects to migrate, then, indicates limits on the ability of states to tax their subjects. Furthermore, it also gives some indication of the choices the subjects face. A larger choice of places to which subjects can migrate will attenuate the ability of governments to overtax, and therefore it encourages subjects to produce more goods.

The next variable discussed is EUROPE, a dichotomous variable equaling one for European countries and their colonial off-springs and zero, otherwise. Two theories support this variable. The first expressed by Joel Mokyr (2017) posits that European culture encouraged scientific endeavor that led to an enhanced adoption of more productive technology even in the period before the industrial revolution. These types of states seemed to have developed in Europe especially its western and northern parts. The populations of four former European colonies, Canada, Australia, New Zealand, and the United States, were made up mostly of emigrants from northern Europe, and they have evolved into European type societies. Consequently, they are classified as European societies for this variable.

The second theory supporting a European distinctiveness has the same reasoning as that behind the NEIGHNO variables (Jones, 2007; and Karayalcin, 2008). It posits a bargaining relationship between each country's government and its population. As stated above, the rulers provide the services from government in exchange for taxes. Accordingly, the subjects pay taxes for the desired services provided by the government rulers. In most situations, the government is a monopoly provider giving it considerable if but not necessarily complete market power. The power of the ruler is constrained by a number of factors. First some states have constitutions that limit their power. These states have less market power than others. Second, the leaders of the states often recognize that a too high price for state services can result in less revenue; thus, some situations call for restraint in taxes (Van Zanden & Ma, 2017). This situation would seem to have a greater presence in the usual smaller European countries than elsewhere.

Biological research has suggested another hypothesis. It is that the genetic composition of populations has affected economic growth in the last 10,000 years. Ashraf and Galor (2013) finds a connection between the genetic diversity of a nation's population and GDP per capita. Based on several theoretical papers, Ashraf and Galor posits that up to a point genetic diversity in a society leads to a population having a greater ability to use their intelligence to adopt their economy to the environment thereby increasing the society's productivity. After a given threshold, however, the greater diversity leads to social conflicts which lessen the ability of a society to produce goods and services with a given resource capacity. Consequently a curvilinear relationship exists between the genetic diversity of a society and its productivity. Essentially as a population becomes more genetically diverse GDP per capita rises up to a point. At a certain threshold, however, the relationship becomes negative. Thus, genetic diversity has a curvilinear relationship to economic success. It has been found that this genetic diversity is negatively correlated with the distance the population and its ancestors have traveled from the eastern African origin of the human species. The sub-Sahara African populations have the most diversity, and the Pacific and western hemisphere populations have the least.

Consequently, the wealthiest countries tend to be in Eurasia with middling amounts of genetic diversity. Ashraf and Galor (2013) exploits this relationship to posit a curvilinear relationship between economic success and genetic diversity. Here, I use the relationship to analyze the impact of genetics on economic performance. To obtain this statistic, I measure the distance of each observation country from Addis Ababa, Ethiopia by the circuitous route that their ancestors hypothetically took from Ethiopia. Essentially all the populations in the sample are hypothesized to have gone from Addis Ababa to Cairo, Egypt and then spread out over the globe by a set of different routes.

The populations in North Africa and the Middle East all the way to India are assumed to have migrated from Ethiopia to Cairo and then from Cairo to each of their population centers. The populations of European countries are assumed after coming from Ethiopia to have migrated from Cairo to Istanbul and then their eventual countries. The populations of East Asia and Oceania are assumed (after arriving in Cairo) to have migrated from Egypt to Phnom Penh, Cambodia and then to their eventual population center. The populations (pre-Columbian) of the western hemisphere are hypothesized to have gone from Cairo to Anadyr, Siberia and then to Prince Rupert, Canada and hence to their final destination population center.

To depict this situation a variable, called GENETIC, is included in the model. It equals the total distance in miles between Addis Ababa and the population center of each observation country by the routes described above. For France, GENETIC would be the distance from Addis Ababa to Cairo plus the distance from Cairo to Istanbul plus the distance from Istanbul to Paris. For China, GENETIC would be the distance from Addis Ababa to Cairo plus the distance from Cairo to Phnom Penh plus the distance from *that* city to Beijing.

Thus, a negative but quadratic relationship should exist between GDPD and GENETIC as depicted by the distance from Addis Ababa. The sample consists only of countries for which GDP and population data are available for the years 1 BCE to 1850. (See Appendix 2 for list of countries.) For only one country south of the Sahara Desert are all the necessary data available. Ashraf and Galor (2013) have hypothesized a nonlinear negative relationship between, GDPC and GENETIC. This would call for a quadratic variable, the square of GENETIC, GENETIC2. Thus, the curvilinear nature of this relationship can be depicted by positing the coefficient for GENETIC to be negative and that for GENETIC2 to be positive.

Finally, a theory has been developed that posits a curvilinear relationship between GDPC and the distance of country from the technological leading nation in the world. If a nation was close to the technological leader, it could better use the leader's knowledge than could more distant countries. But the countries farthest away became more self-reliant, and consequently they may have advantages over some countries closer to the leader (Ozak, 2018). The idea behind this theory is countries closer to the technological leader can appropriate its technology sooner than other nations. The countries farthest away are the most self-sufficient, and they may be able to learn and invent better methods than some of the somewhat nearer countries.

This implies a curvilinear relationship between the GDPC of a country and its distance from the world technological leader. Therefore, the variables, TECHI, equaling the distance between the observation country and the technological leader and TECHI2, the square of TECHI, are added to the model. The identity of the technological leader changed over the sample. For the year, 1CE, the technological leader was Italy; for 1000, it was Iraq, the center of the Moslem Caliphate; for 1500, it was again Italy; for 1600 and 1700, it was the Netherlands, and for 1820 and 1850, it was Great Britain and after those dates, it was the United States.

In summary, this paper initially estimates this specification models with the following variables:

GDPC=f (AREA, LAT, ELEV, LANDL, COL1, COLO, MAREA, EFFL, EFFL1, NEIGHNO, EUROPE, GENETIC, GENETIC2, TECHI, TECHI2, YEAR Dummies).

All specifications of the model will contain the first four terms along with different combinations of the other X variables. See Appendix II for the statistics and characteristics of these variables.

Results for the Classical Model Estimation

In this section, the results for the classical estimation method of the model are given. In the subsequent section, the results of the Bayesian analysis are discussed, and then comparisons are made. To put this in context, it must be recognized that that I use here two empirical estimation philosophies: the classical and Bayesian. With the former, I estimate a model that includes all the variables thought to affect GDP per capita. The Bayesian approach is comprised of two techniques: Extreme Bounds Analysis (EBA) and Bayesian Analysis of Classical Estimation (BACE). Below I first describe and discuss the classical estimating results, presenting first the estimation of the whole model and then the results for a model using only the statistically significant or close to significant variables in that first model.

In Table I, the results for the classical technique are displayed. For the whole model, on the left side of Table I, the results are reasonable from a statistical viewpoint. For this model, the adjusted R^2 is 0.547 inferring that much of the variation in GDPC is accounted for by the model. The F value, 29.57, implies a high likelihood that model explains much of the variation in GDPC, the dependent variable. The coefficient for AREA is insignificant implying that area does not affect GDPC in this sample. LAT is significant implying a positive relationship between the distance from the equator of the observation country and its GDPC. The coefficient for ELEV is negative and significant implying that GDPC falls as a country's elevation rises. COL1 is statistically significant at a positive value, while COLO is significant at a negative value implying that having colonies raises GDPC and being a colony lowers it. These last results are consistent with earlier work (Rogers, 2019).

The results for the variable, MAREA, are insignificant, but they have the predicted sign. Thus, it may very well be that the industrial revolution did not change the relationship between the size of physical resources of a country and its GDPC or that this variable is not a good indicator of the Malthusian effects. While the results do not conclusively confirm the hypothesis, they do present some evidence supporting it. Thus, it is retained in the second attenuated model.

The variable, EFFL, depicts the presence of efflorescence as hypothesized by Goldstone (2002). That EFFL has a positive and significant sign implies EFFL raises GDPC backing the Goldstone hypothesis. The regression coefficient for the variable, EFFL1, however, is negative implying the state of efflorescence before the Industrial Revolution led to lower GDPC other things equal. Thus, the results are partially contrary to the hypothesis of this paper. The variable, NEIGH1, has a positive sign consistent with the hypothesis of this paper, but it is not statistically significant. The dichotomous variable, EUROPE, is negative – inconsistent with this paper's hypothesis.

Due to the hypothesis of negative correlation between genetic complexity and the distance from the East African home of the human race, the positive coefficient for GENETIC implies that lower genetic diversity raises income, the opposite of our posited hypothesis. It is statistically significant but the variable for a curvilinear relationship, GENETIC2, equaling GENETIC squared, is not significant.

The posited relationship between GDPC and the distance from the technologically leading nation is supported by the results of this model. The coefficient for TECHI, the distance of the observation nation from the world technological leader, is statistically significant in the predicted negative direction. Furthermore, the coefficient of TECHI2, the square of TECHI, is positive and statistically significant. This supports the hypothesis of non-linearity for this variable.

The dummy variables depicting the different years are all negative. This result is consistent with the positive trend in GDPC over the sample from 0CBC to 1940. For the later years, however, the coefficients are not statistically significant. In general, it indicates that on the whole the other variables in the model do not fully explain the growth of GDPC.

Thus, the classical results support some of the above posited hypotheses: on elevation, colonialization, Efflorescence for modern countries, technology, and time, but it does not support some of the other hypotheses. In general, it indicates that on the whole the other variables in the model do not fully explain the growth of GDPC in the sample period.

When the GDPC model is estimated without most of the variables with insignificant signs, the result is very similar to those of the full model. The adjusted R^2 is 0.549 a little higher than that of the full model. The F value implies that there is a very small likelihood that the variables in the model are not connected to the dependent variable, GDPC. All the variables that were not eliminated are significant at the 95th per cent level except three, MAREA, GENETIC, and GENETIC2.

In total, the classical results indicate that most of the included variables impact GDPC in the way expected by theory. However, they do not provide a full explanation for the variation in GDPC, given adjusted R^2 values of 0.547 and 0.548 indicating that there is still much variation to be explained. This result leads to the possibility that using another approach could yield more information and conclusions. Thus, I now examine the Bayesian method results.

The Results for the Bayesian Analysis Estimation

Here I explain and discuss the results of Bayesian models. To simplify matters, I perform the Bayesian analysis only on the second model displayed in Table I. While this procedure is less inclusive than using the first model, it is still complex making for the estimation of over 1000 models (1024 to be exact). Thus, I only estimate the Bayesian model for variables in the second (right side) attenuated model in Table I.

I now examine the application of the two Bayesian techniques to the variables for which there were significant or close to significant results for the classical analysis. The particular techniques are Extreme Bounds Analysis (EBA) and Bayesian Analysis of Classical Estimation (BACE). In Table II, the results of these two approaches are presented.

In the Extreme Bounds Analysis (EBA) analysis, I estimate every combination of the ten variables included in the Bayesian analysis. Each model estimated in the EBA analysis combines the ten variables on which Bayesian analysis is focused in the possible 1024 different models. Table II displays the results for both the EBA and BACE approaches. Starting with Extreme Bounds Analysis, I find that for four variables the approach unambiguously supports the original hypotheses. They are COL1, MAREA, EFFL, and TECHI2. For these variables, the Extreme Bounds Analysis supports the paper’s hypotheses in that all the regression coefficients in the over 500 models in which each variable is included have the signs predicted by the paper. While a Bayesian approach implies estimating 1024 models, any one of the Bayesian variables appears in only 512 of the estimation equations. For the last variable, TECHI2, the results are ambiguous because the variable, TECHI, which theory combines with this variable does not always have the predicted sign for all the specifications.

For the variables, COLO, EFFL1, GENETIC2, and TECHI, the results are mixed; the coefficients in some equations are as predicted, and in some others, they are not. Therefore, the EBA offers no robust results for those variables in that some models show one direction for the variable influence and others show another. Consequently, the EBA results for these varaibles are ambiguous.

For the first of these variables, COLO, the mixed results, however, are still suggestive if not conclusive. For 506 (98.8%) of the models in which it appears, it has the sign predicted by the theory implying that colonies were exploited by their mother countries. While the result is highly suggestive, it does not pass the criterion usually employed in Extreme Bounds Analysis; thus, it is not fully robust.

For the variables, EUROPE and GENETIC, the EBA results were robust, but contrary to the hypothesized theory. For EUROPE, the robust models indicate that other things equal, being European depressed GDPC. Perhaps, other variables raised the expected income of Europe from its normal height so much it obscured the underlying negative conditions that kept its expected income so low. For GENETIC, all model have signs unpredicted by theory.

The second Bayesian technique, BACE or Bayesian Analysis of Classical Estimation, seems to better support the veracity of the classical estimation technology. Essentially the Bayesian Analysis of Classical Estimation calculates a weighted average of the coefficients for the Bayesian models. (See Sala-i-Martin, Doppelhofer, and Miller, 2004.) The weights are computed from the standard errors of the coefficients for each of the models calculated for each rendition of the model. The models consist of all the possible combinations of the Bayesian variables. These variables, then, consist of all of the variables in my model except for LAT, ELEV, and the year dummies**,** which are included in all of the 1024 estimated models. For the rest of the variables, each model includes a combination of the ten other variables. These ten different variables are combined in all possible different ways which lead to 2^{10} or 1024 different ways. The Bayesian Analysis of Classical Estimation sets up for each variable a weighted average of the model regression coefficients and coefficient standard errors. From the ratio of these two averages, a t value can be computed from which statistical inferences can be made on this paper's hypotheses.

The right-most column in Table II gives these weighted coefficients. Under each coefficient is its t value computed from the ratios of the weighted coefficients and standard errors. From these ratios, inferences can be made on the likelihood of the variable having a given direction. The results for the BACE analysis seem a little more consistent with classical models than with the EBA method. For six of the ten variables, the results are statistically significant in the expected direction. For COL1, the coefficient is positive and significant at the 99 per cent level implying colonialist countries had higher GDPC. The coefficient estimate for COLO is negative and significant indicating that having been a colony reduces GDPC. These results are consistent with the hypotheses of this paper and past work (Rogers 2019).

For four other variables, the BACE findings are consistent with the classical results. The negative and significant coefficient for MAREA, the variable reflecting the Malthusian condition prior to the industrial revolution led to lower national income. The BACE results for the variable, EFFL, also support the hypothesis described above in that conditions in some countries led to higher income where the condition of efflorescence seem to prevail. Unlike the results for the EBA, the quadratic model for technology has the expected signs, and they are statistically significant.

For four variables, EFF1, EUROPE, GENETIC, and GENETIC2, the BACE findings contradict the original hypotheses of this paper. Consistent with the classical and EBA approaches, they find that EFF1 and EUROPE have negative and significant signs implying that being efflorescent before the industrial revolution and being in Europe lowered GDPC other things equal. Thus, further examination is needed on the impacts of being efflorescent before the industrial revolution and being European.

The BACE results, then, lend some support to the classical results shown in Table 1. At least some of the classical results are supported by both the EBA and BACE results, and furthermore, the unexpected results for EFFL1 and EUROPE are consistent between not only the two Bayesian results but also between the Bayesian models and the classical approach.

Conclusion

This paper applies Bayesian methods to ascertain whether the results from classical econometric model stand up under a variety of different model specifications. Thus, the Bayesian methods are used to check the sensitivity of these results to variation in model specification.

Some variation is observed in the results of the Bayesian models. While the Bayesian Analysis of Classical Estimation (BACE) is consistent with most results from the classical approach, the findings for the Extreme Bounds Analysis (EBA) often contradict the classical approach results. Essentially the Bayesian methods support the hypotheses on Colonies, Efflorescence, and the Malthus impact. The Bayesian results for the other variables are not robust.

The BACE approach does lend some support to the use of classical modeling, but the EBA leads to caution in stopping with a classical model and ignoring the possibility of further investigation exposing given classical models to contradiction due to undetected influences. Thus, future econometric research on long term GDPC should combine classical modeling with Bayesian methods as this paper has done.

Table 1
Regressions for Two GDPC Models for Sample Countries for the Period between 1 BCE and 1940

Variable Name	Coefficients for Full Linear Model for all Variables	t Value	Coefficients for the Model with Significant Coefficients Variables	t Value
(Intercept)	1613.00		1579.000	
AREA	0.00034	0.692		
LAT	40.7500	6.951 **	41.360	7.376 ***
ELEV	-0.06110	-2.168 *	-0.05564	-2.042 *
LANDL	170.000	0.865		
COL1	593.100	3.627 ***	588.100	3.707 ***
COLO	-310.700	-2.044 *	-310.6	-2.078 *
MAREA	-0.00013	-1.903	-0.00010	-1.946
EFFL	1568.00	8.512***	1591.00	8.928***
EFFL1	-1224.00	-2.465 *	-1263.000	-2.565 *
NEIGH1	2.1180	0.155		
EUROPE	-868.200	-3.836 ***	-857.80	-4.373***
GENETIC	0.093	1.965 *	0.08971	1.930
GENETIC2	-0.000002	-1.523	-0.0000024	-1.472
TECHI	-0.4087	-5.816 ***	-0.04143	-5.942 ***
TECHI2	0.000043	6.126 ***	0.000043	6.271 ***
YRO	-1728.000	-4.869 ***	-1683.00	-5.556 ***
YR1000	-1924.000	-5.396 ***	-1863.00	-5.714 ***
YR1500	-1617.000	-4.817 ***	-1563.00	-5.139 ***
YR1600	-1862.000	-5.633 ***	-1811.000	-6.082 ***
YR1700	-1628.000	-5.103 ***	-1573.000	-5.488 ***
YR1820	-1495.000	-5.660 ***	-1433.000	-6.339 ***
YR1850	-2405.000	-8.389 ***	-2337.000	-9.384 ***
YR1870	-1751.000	-7.086 ***	-1686.000	-8.334 ***
YR1900	-1033.000	-3.870 ***	-955.1000	-4.252 ***
YR1913	-483.90	-2.040 *	-408.800	-2.133 *
YR1920	-525.0	-2.087 *	-446.100	-2.137 *
YR1925	-200.6	-0.815		
YR1930	-44.87	-0.183		
YR1935	-74.83	-0.305		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of Observations	687	687
Adjusted R-squared:	0.547	0.549
F-statistic:	29.52	37.36

Table 2

Extreme Bounds Analysis and Bayesian Averaging of Classical Estimates for GDPC Variables

Variable		Maximum Impact	Minimum Impact	Number of Models with Predicted Signs**	BACE Estimate
COL1	Coefficient	1109.315	382.608	512(100.0%)	776.412
	t values	(6.827)	(2.442)		(4.773)
COLO	Coefficient	34.104	-878.654	506(98.8%)	-391.118
	t values	(0.023)	(-5.985)		(-2.559)
MAREA	Coefficient	-0.000008	-0.00023	512(100.0%)	-0.00014
	t values	(-0.151)	(-4.275)		(-2.644)
EFFL	Coefficient	1943.96	1270.60	512(100.0%)	1586.64
	t values	(11.553)	(7.031)		(9.231)
EFFL1	Coefficient	925.304	-1592.194	258(50.2%)	-238.204
	t values	(1.845)	(-3.233)		(-0.470)
EUROPE	Coefficient	-317.938	-1270.154	0(0.0%)	-749.248
	t values	(-1.489)	(-7.031)		(-3.770)
GENETIC	Coefficient	0.2366	0.0144	0(0.0%)	0.1522
	t values	(5.375)	(1.317)		(3.356)
GENETIC2	Coefficient	0.000003	-0.000007	261(51.0%)	-0.000004
	t values	(7.918)	(-4.502)		(-2.341)
TECHI	Coefficient	0.0874	-0.4931	217(42.4%)	-0.3590
	t values	(3.503)	(-7.336)		(-5.449)
TECHI2	Coefficient	0.00005	0.000002	512(100%)	0.00004
	t values	(7.232)	(0.723)		(6.385)

** In parentheses is the percentage of the total with the predicted sign.

References

Auer, R. (2013). Geography, institutions, and the making of comparative development, *Journal of Economic Growth,* (2013) *18*, 179–215.

Ashraf, Q., & Galor, O. (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, *103*(1), 1-46.

Barro, R. J. and X. Sala-i-Martin (2004). *Economic Growth* (2nd Ed.*)*. MIT Press.

Bolt, J., & Van Zanden, J. L. (2014). The Maddison Project: Collaborative research on historical national accounts. *Economic History Review*, *67*(3), 627–651.

Chaudry, A., & Garner, P. (2006). Political competition between countries and economic growth. *Review of Development Economics*, *10*(4), 666–682.

Erdkamp, P. (2016). Economic Growth in the Roman Mediterranean world: An early good-bye to Malthus. *Explorations in Economic History*, *60*, 1-20.

Galor, O., & Moav, O. (2006). Das human-kapital: A theory of the demise of the class structure. *Review of Economic Studies*, *73*(1), 85-117.

Galor, O., Moav, O., & Vollrath, D. (2009). Inequality in landownership, the emergence of human-capital promoting institutions, and the Great Divergence. *Review of Economic Studies*, *76*, 143–179.

Galor, O., & Weil, D. N. (2000). Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. *American Economic Review*, *90*(4), 806-828.

Gassebner, M., J. Gutmann, J., & Voigt, S. (2016). When to expect a coup d'état? An extreme bounds analysis of coup determinants. *Public Choice 169*, 293–313.

Goldstone, J. A. (2002). Efflorescences and economic growth in world history: Rethinking the 'Rise of the West' and the Industrial Revolution. *Journal of World History*, *13*(2), 323-389.

Johnson, N. D., & Koyama, M. (2017). States and economic growth: Capacity and constraints. *Explorations in Economic History,* 1-20.

Jones, E. L. (2007). *Growth Recurring: Economic Change in World History*. University of Michigan Press.

Karayalcin, C. (2008). Divided we stand, united we fall: The Hume-North-Jones Mechanism for the rise of Europe. *International Economic Review*, *49*(3), 973-997.

Levine, R., & Zervos, S. (1993). What we have learned about policy and growth from cross-country regressions? *American Economic Review*, *83*(2), Papers and Proceedings of the Hundred and Fifth Annual Meeting of the American Economic Association, 426-430.

Levine, R., & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review*, *82*(4), 942-963.

McGuire, M. C., & Olson, M. (1996). The economics of autocracy and majority rule: The invisible hand and the use of force. *Journal of Economic Literature*, *34*(1), 72-96.

Maddison, A. (2001). *The world economy: A millennial perspective*, Development Centre of the Organization for Economic Co–Operation and Development.

Malthus, T. R. (1976). *An essay on the principle of population: Text, sources and background, criticism.* (P. Appleman,Ed.). Norton. (Original work published 1798)

Mokyr, J. (2017). *A culture of growth: The origins of the modern economy*. Princeton University Press.

Owen, A. L., Videras, J., & Davis, L. (2009). Do all countries follow the same growth process? *Journal of Economic Growth*, *14*, 265–286.

Ozak, O. (2018). Distance to the pre-industrial technological frontier and economic development. *Journal of Economic Growth*, *23*, 175–221.

Moreno, R. & Trehan, B. (1997). Location and the growth of nations. *Journal of Economic Growth*, *2*, 399–418.

Rogers, R. P. (2019). The impact of colonization on home country wealth. *Journal of Business Diversity* 19(4), 60-76.

Sala-i-Martin, X. (1997). I just ran two million regressions. *American Economic Review*, *87*(2), Papers and Proceedings, 178-183.

Sala-i-Martin, X., Doppelhofer, G., & Miller, R. (2004). Determinants of long-term growth: A Bayesian Averaging of Classical Estimates (BACE) approach. *American Economic Review*, *94*(4), 813-835.

Sng, T. (2014). Size and dynastic decline: the principal-agent problem in late imperial China, 1700–1850. *Explorations in Economic History*, *54*, 107–127.

Solow, R. M. *(1957).* Technical change and the aggregate production function," *Review of Economics and Statistics*, *39*(3), 312–20.

Solow, R. M. *(1994).* Perspectives on growth theory. *Journal of Economic Perspectives*, *8*(1), 45-54.

Sturm, J., & Haan, J. (2005). Determinants of long-term growth: New results applying robust estimation and extreme bounds analysis. *Empirical Economics*, *30*,597–61.

Van Zanden, J., & Ma, D. (2017). *What makes Maddison right?* Maddison-Project Working Paper WP-7.

Van Zanden, J., & van Leeuwan, B. (2012). Persistent but not consistent: The growth of national income in Holland 1347–1807. *Explorations in Economic History*, 49, 119–130.

Appendix 1
List of Countries in the Sample

Number	Country
1.	Albania
2.	Argentina
3.	Australia
4.	Austria
5.	Belgium
6.	Bulgaria
7.	Bolivia
8.	Brazil
9.	Canada
10.	Switzerland
11.	Chile
12.	China
13.	Colombia
14.	Costa Rica
15.	Czechoslovakia
16.	Cuba
17.	Germany
18.	Algeria
19.	Ecuador
20.	Egypt
21.	Spain
22.	Finland
23.	France
24.	United Kingdom
25.	Greece
26.	Guatemala
27.	Honduras
28.	Hungary
29.	Indonesia
30.	India
31.	Ireland
32.	Iran
33.	Iraq
34.	Italy
35.	Jamaica
36.	Jordan
37.	Japan
38.	S Korea
39.	Lebanon
40.	Sri Lanka
41.	Morocco
42.	Mexico
43.	Myanmar
44.	Malaysia

Appendix 1
List of Countries in the Sample (Continued)

45.	Nicaragua
46.	Netherlands
47.	Norway
48.	Nepal
49.	New Zealand
50.	Panama
51.	Peru
52.	Philippines
53.	Poland
54.	Portugal
55.	Romania
56.	Saudi Arabia
57.	Singapore
58.	El Salvador
59.	FUSSR
60.	Sweden
61.	Syria
62.	Thailand
63.	Tunisia
64.	Turkey
65.	Taiwan
66.	Uruguay
67.	United States
68.	Venezuela
69.	Viet Nam
70.	F Yugoslavia
71.	South Africa

Appendix 2
Mean and Standard of the Variables in the GDPC Models.

Variable Name	Mean	Standard Deviation
GDPC (GDP per capita)	2255.509	1992.337
POPULATION (population of observation country)	22056.84	60071.22
AREA (area of observation country)	714416.7	1992.337
LAT (Latitude of observation country)	35.223	16.9670
ELEV(Latitude of observation country)	1881.705	2041.513
LANDL (dummy variable for landlocked country)	0.088792	0.28465
COL1(dummy variable for landlocked colonist country)	0.175509	0.37871
COLO (dummy variable for country being a colony)	0.247787	0.431846
MAREA (variable indicating the Malthus effect)	356480.8	1103272
EFFL (dummy for Efflorescent whole sample)	0.419214	0.49379
EFFL1 (dummy for Efflorescent prior to 1850)	0.016012	0.125611
NEIGHNO	9.740902	6.790265
EUROPE	0.43377	0.495955
GENETIC	11249.405	7249.8434
GENETIC2	179032827	205529522
TECHI	3953.163	2713.156
TECHI2	22978000	27297443