Economic Inquiry

WANT EXPORT DIVERSIFICATION? EDUCATE THE KIDS FIRST

MICHAEL JETTER and ANDRÉS RAMÍREZ HASSAN

This paper uses Bayesian model averaging to uncover the true determinants of export diversification among 36 potential factors, and thus 2^{36} potential models. Using data from 2001 to 2010, our results reveal two strong predictors: Primary school enrollment (99.7% posterior inclusion probability in the true model) raises export diversification, whereas the share of natural resources in gross domestic product (98.6%) lowers diversification levels. The importance of basic education coverage offers policymakers an opening toward diversifying exports, at least in the long run. This result is robust to accounting for the endogeneity of income levels by applying an instrumental variable BMA method. (*JEL* C11, F1, F34, O1, O2, O11)

I. INTRODUCTION

As globalization has taken on unforeseen dimensions, various aspects of international trade are gaining importance. Economists and politicians do not discuss only pure trade openness anymore, they also discuss the form and diversity of a country's export basket. Influential works by Imbs and Wacziarg (2003) and Cadot, Carrère, and Strauss-Kahn (2011) describe the pattern of export diversification over development levels of countries. In fact, a diversified export basket seems to foster economic growth in several ways, as discussed in more detail below. However, the roots of export diversification have received less attention. But if diversifying exports carries benefits in the development of an economy, then it is useful to understand what determines diversification levels in the first place. For example, if the real determinants lie in the range of policies (such as trade regulations or education) then the degree of diversification can be influenced. But if country-specific conditions (such as geographical location, population size, or colony status) play a dominant role, little remains to be done for policymakers. This paper constitutes a step toward understanding the causes of export diversification by systematically analyzing 36 potential determinants.

The potential benefits from export diversification are suggested to be substantial. For instance, Jansen (2004), Agosin (2006), and Agosin (2009) find that diversifying exports alleviates volatility in macroeconomic variables, which in turn supports economic growth. Other arguments consist of learning-by-exporting and knowledge spillovers (Hausmann, Hwang, and Rodrik 2007; Herzer and Nowak-Lehnmann 2006). Overall, the existing literature consistently suggests beneficial growth consequences from diversifying exports (see Al-Marhubi 2000; Cadot, Carrère, and Strauss-Kahn 2012; Lederman and Maloney 2003; Newfarmer, Shaw, and Walkenhorst 2009).[1] Although there exists a consensus about the benefits of export diversification for economic growth, the literature on the

Jetter: School of Economics and Finance, Universidad EAFIT, Carrera 49 7 Sur 50, Medellin, Antioquia 3300, Colombia; Business School, University of Western Australia, Perth, Australia (July 2015 onwards). Phone 0057 317 433 7003, Fax 0057 054 266 4284, E-mail mjetter7@gmail.com

Ramírez Hassan: School of Economics and Finance, Universidad EAFIT, Carrera 49 7 Sur 50, Medellin, Antioquia 3300, Colombia. Phone 0057 312 398 5240, Fax 0057 054 266 4284, E-mail aramir21@eafit.edu.co

1. For country- and region-specific analyses one might consider Alwang and Siegel (1994), Amin Gutiérrez de Piñeres and Ferrantino (1997), Taylor (2003), Petersson (2005), Beine and Coulombe (2007), or Volpe Martincus and Carballo (2008).

ABBREVIATIONS

BMA: Bayesian Model Averaging
GDP: Gross Domestic Product
HHI: Herfindahl–Hirschman Index
IVBMA: Instrumental Variable Bayesian Model Averaging
MCMC: Markov Chain Monte Carlo
OPEC: Organization of the Petroleum Exporting Countries
PIP: Posterior Inclusion Probability
PMP: Posterior Model Probability

Economic Inquiry
(ISSN 0095-2583)
Vol. 53, No. 4, October 2015, 1765–1782

doi:10.1111/ecin.12213
Online Early publication March 9, 2015

roots of export diversification is less consistent. Previous works suggest various potential factors including geographical components, schooling, income levels, or attitudes toward international trade. Table 1 summarizes the most prominent results, but at the same time reveals persistent problems that have plagued the interpretational power of these analyses. Consider the number of included regressors, which varies between 7 and 23 (not counting the nonparametric paper). Only two variables appear significant in more than one study, and it remains unclear which variables should form part of a standard regression estimating export diversification. In general, theories about the determinants of export diversification do not necessarily exclude each other. Thus, model uncertainty seems to be a major problem. Beyond that, the usual econometric limitations aggravate the development of a profound long-term analysis. These limitations include scarce data, a need to account for business cycles, and several forms of endogeneity, such as measurement error or potential reverse causality. A major problem is posed by income levels, as it becomes hard to identify the true direction of causality: richer countries may be able to diversify more, but export diversification might itself lead to higher income levels.

This paper uses Bayesian model averaging (BMA) and instrumental variable Bayesian model averaging (IVBMA) to reveal the true model determining export diversification. The strength of the BMA method lies precisely in finding the true model when faced with a small sample size and numerous candidates for explanatory variables. We analyze 36 potential long-run determinants of export diversification in one cross-sectional sample of 105 countries, sorting through all possible combinations of variables (2^{36}). Averaging variables over a decade minimizes the impact of business cycles, temporary exogenous shocks, and measurement error. Furthermore, using a recently developed IVBMA method, we are able to address the latent reverse causality problem between diversification and income levels.

The following section presents our methodology, followed by a description of the data in Section III. Section IV presents our results, followed by the conclusion in Section V.

II. METHODOLOGY

Tracing the long-run determinants of export diversification poses several problems. First, one

TABLE 1
Recent Empirical Studies on the Determinants of Export Diversification

Paper	Authors	Dependent Variable	# of Regressors	Empirical Method	Significant Determinants
Determinants of Export Diversification Around the World: 1962–2000	Agosin, Alvarez, and Bravo-Ortega (2012)	HHI, Gini, Theil	7	GMM system estimator	Remoteness (−), trade openness (−), schooling (+), human capital interacted with terms of trade
Trade Facilitation and Export Diversification	Dennis and Shepherd (2011)	# of exports	23	Poisson, IV Poisson, 2 SLS	Costs of export, market entry, and international transport (all −)
Employment and Export Specialization along the Development Path: Some Robust Evidence	Parteka (2010)	HHI, Gini, Theil	2	Semiparam. regression	GDP (+), GDP squared (−)
Determinants of Export Diversification: An Empirical Investigation	Parteka and Tamberi (2008)	Gini, Theil	34	OLS	Population (+), distance from major markets (−), GDP (+), institutional capacity to favor free trade (+)
Explaining Export Diversification: An Empirical Analysis	Bebczuk and Berrettoni (2006)	HHI	8	OLS	Primary products (−), South America (−), Africa (−),
Diversification: Toward a New Paradigm for Africa's Development	Ben Hammouda et al. (2006)	HHI	11	OLS	Gross capital formation (+), gross capital formation squared (−), GDP (+), GDP squared (−), trade openness (−), industrial production (+), inflation (−), exchange rate (−), governance (+), conflict (−)

needs to smooth out business cycles and contain the impact of measurement error, which can be of particular importance in some developing regions, for instance. Thus, one might choose to average variables over 5 or 10 years and thereby end up with fewer observations. Second, as discussed above, the choice of potential explanatory variables is open-ended, implying model uncertainty. Combining these problems produces both a relatively small number of observations and a large number of potential explanatory variables. The BMA methodology is exactly designed to address model uncertainty in these kinds of environments (Leamer 1978; Raftery 1995). Intuitively, it considers all possible combinations of independent variables (here 2^{36}) and then produces a probability of inclusion into the "true" model for every variable (the Posterior Inclusion Probability, PIP).

A. Addressing Model Uncertainty

We start by applying BMA to the generic regression framework estimating export diversification, as measured by the Herfindahl–Hirschman index (HHI). Although the long-term trend of a country's degree of export diversification is relatively stable, the data can exhibit substantial year-to-year variation for numerous countries. This leads us to believe that either (1) short-term fluctuations in export diversification levels and in macroeconomic variables severely affect the short-run value of diversification or (2) the data exhibits measurement error. Averaging variables over 10 years alleviates both problems, and our estimation for country i becomes

$$\text{(1)} \quad \overline{HHI}_i^{2001-2010} = \alpha + \vec{\beta}\,\overline{\vec{X}}_i^{2001-2010} + \varepsilon_i,$$

where $\overline{\vec{X}}_i^{2001-2010}$ contains all 36 potential determinants, which are discussed in detail below. Our main sample consists of 105 countries for which we have data for all variables. Section III.F provides details. The BMA procedure proves especially powerful in cases when the researcher is faced with a large set of potential explanatory variables (in this case 36), but only a relatively limited number of observations (in this case 105 countries). In this context, Raftery (1995) has shown that standard variable selection procedures can lead to misleading results.

In general, the BMA technique is firmly grounded in statistical theory following the rules of probability. BMA minimizes the sum of Type I and Type II error probabilities; its posterior point estimates minimize the mean square error, and its posterior predictive distributions perform better relative to other estimators (Raftery and Zheng 2003). Additionally, there exists a huge difference between the BMA method and other model selection techniques, such as the Bayesian Information Criterion, as the latter method chooses just one model, whereas the former averages over all possible models. This difference makes BMA a stronger statistical tool. Although BMA enjoys a long tradition in statistics (Leamer 1978), its application in economics has only recently come into its own. In particular, BMA has recently been applied to further identify the true determinants of economic growth—a literature that has been pioneered by the seminal papers of Fernandez, Ley, and Steel (2001b), Sala-i Martin, Doppelhofer, and Miller (2004), and, most recently, Moral-Benito (2013a) in a panel setting. Other applications in the international macroeconomic context include those of Eicher, Henn, and Papageorgiou (2012), who analyze the determinants of the gravity equation model, and Eicher, Helfman, and Lenkoski (2012), who focus on foreign direct investment. In general, Moral-Benito (2013b) provides a detailed survey on the use of BMA methods in economics.

General BMA Framework. A Bayesian solution to model uncertainty involves averaging over all possible combinations of predictors (Leamer 1978). Raftery, Madigan, and Hoeting (1997) show that, in the presence of predictors' uncertainty, the performance of BMA procedures is superior to any single model selected using frequentist arguments. However, this solution may not be practical in some circumstances, as the number of possible variable combinations increases quickly: for K predictors, the number of possible combinations becomes 2^K. This has led to the development of various algorithms based on the Markov Chain Monte Carlo (MCMC) strategy.[2]

With $\mathscr{M} = \left\{M_1, M_2, \ldots, M_{2^K}\right\}$ denoting the set of considered models, each model depends on a vector of parameters θ_r (with $r = 1, 2, \ldots, 2^K$) characterized by a prior $\pi(\theta_r|M_r)$ and a likelihood $\pi(y|\theta_r, M_r)$, where y is an N-dimensional vector such that $y = \beta_0 i_N + X_r\beta_r + \mu$, i_N is an $N \times 1$ vector of ones, X_r is an $N \times K_r$ matrix of predictors,

2. In addition to the MCMC strategy, the literature has produced other mechanisms to implement BMA, for example Occam's window (Madigan and Raftery 1994).

and μ is a vector of stochastic errors. Additionally, there is the posterior distribution $\pi(\theta_r|y, M_r)$. Using standard probability arguments, the posterior $\pi(\theta_r|y)$ is then given by

$$(2) \quad \pi\left(\theta_r|y\right) = \sum_{r=1}^{2^K} \pi\left(\theta_r|y, M_r\right)\pi\left(M_r|y\right).$$

The BMA logic obtains results for every model under consideration and averages them. Further, the posterior model probability (PMP) is proportional to $\pi(y|M_r)\pi(M_r)$, that is

$$(3) \quad \pi\left(M_r|y\right) \propto \pi\left(y|M_r\right)\pi\left(M_r\right)$$

with a marginal likelihood given by

$$(4) \quad \pi\left(y|M_r\right) = \int_{\mathscr{R}^r} \pi\left(y|\theta_r, M_r\right)\pi\left(\theta_r|M_r\right)d\theta_r$$

and $\pi(M_r)$ being the prior model probability. However, two main issues remain when trying to implement the BMA strategy: the integral in Equation (4) is difficult to implement, and the number of variable combinations can be enormous. These issues can be handled using a birth/death MCMC algorithm—an adaptation from the mechanism originally developed by Madigan, York, and Allard (1995).

Practical BMA Issues. The birth/death MCMC algorithm consists of a mechanism for sampling over a model space $\mathscr{M}$, based on a Metropolis–Hastings algorithm (Hastings 1970; Metropolis et al. 1953). It simulates a chain of models, denoted by $M^{(s)}$ (for $s = 1, 2, \ldots, S$), where the algorithm draws candidate models from a particular distribution over the model space and then accepts them with a certain probability. If a candidate model is not accepted, the chain remains in the current model (Koop, Poirier, and Tobias 2007). A candidate model M^c is drawn randomly from the set of models, including (1) the current model $M^{(s-1)}$, (2) all models which delete one predictor from $M^{(s-1)}$, and (3) all models which add one predictor to $M^{(s-1)}$. The acceptance probability then becomes

$$(5) \quad \alpha\left(M^{(s-1)}, M^c\right) = Min\left\{\frac{\pi\left(y|M^c\right)\pi\left(M^c\right)}{\pi\left(y|M^{(s-1)}\right)\pi\left(M^{(s-1)}\right)}, 1\right\}.$$

In general, BMA comes with an immense computational burden. As a consequence, the literature favors the so-called natural conjugate approach—in this case the normal linear model, where $\mu \sim \mathcal{N}_N\left(0, \tau^{-1}I_N\right)$. This leads to $\pi\left(\beta_r|\tau\right) \sim \mathcal{N}_{K_r}\left(0_{K_r}, \tau^{-1}\left(g_r\left(X_r'X_r\right)\right)^{-1}\right)$. In practice, this means that we center priors over the hypothesis that explanatory variables do not have an effect on the dependent variable, and that the covariance of the explanatory variables is proportional to the comparable data-based quantity. Additionally, we assume a non-informative prior for common parameters to all models, that is, τ and β_0. Specifically, $\pi(\tau) \propto 1/\tau$ and $\pi(\beta_0) \propto 1$. In this context, Fernandez, Ley, and Steel (2001a) recommend selecting $g_r = 1/K^2$, if $N \leq K^2$ or $g_r = 1/N$, if $N > K^2$ after extensive simulation exercises. Ley and Steel (2009), in contrast, propose a Beta-Binomial prior on $\pi(M_r)$, because the resulting prior model distribution is considerably less tight and should thus reduce the risk of unintended consequences from imposing a particular prior model size. The latter method only requires one to choose the prior expected model size. In particular, if the prior expected model size is equal to $K/2$, the model prior is completely flat over model sizes.

With these considerations in mind, we estimate multiple Bayesian normal linear models whose differences are given by the combination of predictors, and where we choose the best models using a birth/death MCMC mechanism based on a Metropolis–Hastings algorithm. Finally, we average these models, obtaining posterior parameters and predictive distributions. Specifically, one model is randomly drawn and accepted if its marginal likelihood is superior to the marginal likelihood of the current model—if not, the new model is randomly accepted according to a probability that depends on the ratio of marginal likelihoods. The procedure is performed many times and eventually we average all possible models weighted by the number of times the specific model was selected. In the end, a variable becomes a candidate to explain the dependent variable if it appears many times in models with high marginal likelihoods.

B. *Considering Endogeneity*

Beyond the problem of model uncertainty, we also extend our analysis by addressing the latent endogeneity problem of income levels in estimating the HHI_i. In particular, we extend our BMA framework to an IVBMA analysis (Karl and Lenkoski 2012; Koop, León-Gonzalez, and Strachan 2012). We follow the procedure developed by Karl and Lenkoski (2012), in which the conditional Bayes factor comparing two models is used within a Gibbs sampler to move between models in a two-stage procedure. This procedure is similar in spirit to a two-stage least squares estimator. In particular, there are two sets of equations: in the first stage, the endogenous variable (gross domestic product [GDP] per capita in our case) is estimated as a function of a set of instrumental variables and the exogenous variables from the principal regression of interest (in this case the HHI). Endogeneity is considered in the covariance matrix of the stochastic perturbations, which relates the first and second stages. The idea is to use the conditional Bayes factor to select the best models in each stage in an iterative way, using a Gibbs sampler. In our case, we use geographical, cultural, and colonial variables as instruments, which will be further discussed in Section IV.D. The second stage then uses the predicted values of GDP per capita (*loggdpcap*) with the remaining exogenous variables in the initial regression framework estimating the HHI.

III. VARIABLES

This section discusses our measurement for export diversification and its potential determinants. Given the large number of potential factors, we only provide a brief explanation for the inclusion of a variable and then refer to the respective papers for further details. Further, we briefly discuss descriptive statistics of our data; data sources are displayed in Table 3.

We sort all explanatory variables in four broad categories: political, macroeconomic, cultural, and geographical factors. Although some variables may well form part of another category, the general idea of sorting is to identify the degree to which a variable is accessible by governments or the private sector. Political and macroeconomic factors are most receptive to policy changes, cultural aspects to a lesser degree, and geographical conditions of a country are mostly impossible to influence. Thus, depending on which variables have significant effects on export diversification, we get an idea as to what extent the level of diversification is potentially modifiable by policies.

A. The HHI Index

We choose a classic representation of export diversification as our dependent variable: the HHI. By using the HHI as the measurement of export diversification, we follow the vast majority of previous analyses (Agosin, Alvarez, and Bravo-Ortega 2012; Cadot, Carrère, and Strauss-Kahn 2011; Imbs and Wacziarg 2003; Parteka 2010). The HHI allows us to capture both the intensive and the extensive margin of diversification.[3]

The raw data for calculating the HHI come from the United Nation's Commodity Trade Statistics Database (ComTrade), permitting us to use annual exports on the 6-digit level of disaggregation for the years 2001–2010.[4] Assuming m goods, a total of US\$ X_{it} of country i's exports in year t, and exporting US\$ x_{ij} of good j, we use the common formula

$$\text{(6)} \qquad \mathrm{HHI}_{it} = \sum_{j=1}^{m} \left(\frac{x_{ijt}}{X_{it}} \right)^2$$

to calculate country i's level of export diversification in year t. This procedure gives us a yearly national diversification index for 105 countries over the time frame 2001–2010. Averaging for every country then gives us the dependent variable:

$$\text{(7)} \qquad \overline{\mathrm{HHI}}_i^{2001-2010} = \sum_{t=2001}^{2010} \left(\frac{\mathrm{HHI}_{it}}{10} \right).$$

Notice that higher values of the HHI signal a higher level of concentration.

B. Political Factors

We start with variables describing the political environment of an economy. First, we include a country's level of democracy, plus the size and efficiency of government. The *polity* variable measures the level of democracy, ranging from −10 (strongly autocratic) to +10 (strongly democratic). In addition, government size (*gov*) is measured as government share of real GDP per

3. This is in contrast to Dennis and Shepherd (2011), who focus on the number of products exported (extensive margin). For the distinction between diversifying along the intensive or extensive margin, one might consider Amurgo-Pacheco (2008) or Cadot, Carrère, and Strauss-Kahn (2011) for empirical analyses. Chaney (2008) provides a deeper theoretical foundation for the intensive and the extensive margin of international trade.

4. Agosin, Alvarez, and Bravo-Ortega (2012) use three-digit level data. Data for the years before 2001 cover substantially fewer countries.

capita.[5] Finally, governance effectiveness (*goveff*) ranges from −2.5 (weak) to +2.5 (strong). Both government size and effectiveness are included in the studies of Ben Hammouda et al. (2006), Dennis and Shepherd (2011), and Parteka and Tamberi (2011).

Addressing trade-related policies, we use two measures: (1) *tradefreedom* ranging from 0 to 100 and (2) the inflation rate (*infl*). The Index of Economic Freedom defines *tradefreedom* as "…the citizen's ability to interact freely as buyer or seller in the international marketplace," addressing any form of regulatory barriers to trade. The inflation rate is included to capture the business environment and financial stability of a country. In additional estimations, we also included three types of tariff rates (weighted mean of all products, manufactured products, and primary products) and the volatility of the inflation rate from 2001 to 2010. However, none of these variables ever come close to being a considerable determinant with posterior inclusion probabilities below 1%.

Another branch of variables considers a country's education system. Various studies propose schooling to play a substantial role in export diversification, e.g., Mengistu (2009), Parteka and Tamberi (2011), or Agosin, Alvarez, and Bravo-Ortega (2012). To this end, we include five education variables: the duration of primary and secondary education, secondary and tertiary enrollment rates (gross percentage), and total primary enrollment (net percentage).[6] Including measurements for both time of education and enrollment rates at various stages allows us to analyze the importance of distinct human capital variables in more detail—an advantage owed to the BMA technique, which does not punish for the inclusion of additional variables. For instance, Agosin, Alvarez, and Bravo-Ortega (2012) employ average years of schooling from Barro and Lee (2000), whereas we distinguish between the duration of primary and secondary education and include enrollment rates at different stages.

Finally, given the extraordinary role of oil in the world economy and the predominant role of the Organization of the Petroleum Exporting Countries (OPEC) in this market, especially toward the end of the twentieth century, we add a dummy variable for OPEC member countries.

C. Macroeconomic Factors

In terms of general macroeconomic conditions, we include *exports* and *imports*, total natural resource rents (*natres*), gross capital formation (*capital*), and foreign direct investment (*fdi*), all measured as a percentage of GDP.[7] All these variables are closely related to international trade and have been analyzed as potential determinants of export diversification (e.g., see Agosin, Alvarez, and Bravo-Ortega 2012; Tadesse and Shukralla 2013). We include *natres* as a proxy for the importance of primary resources in the domestic economy, defined as the sum of oil, natural gas, coal (hard and soft), mineral, and forest rents (World Bank). This variable is particularly important given the discussion about the "natural resource-curse" in the context of export diversification and economic growth (Sachs and Warner 2001, or more recently Gylfason 2009 and van der Ploeg 2011). In addition, we include the percentage of paved roads (roads) in a country as a measurement for infrastructure and transportation systems.

Further, we add the logarithm of GDP per capita (*loggdpcap*), a net barter terms of trade index (year 2000 = 100, *tot*), and fuel exports (as percent of merchandise exports, *fuel*). Either one or a combination of these variables is included in Imbs and Wacziarg (2003), Bebczuk and Berrettoni (2006), Ben Hammouda et al. (2006), Dennis and Shepherd (2011), Parteka and Tamberi (2011), Agosin, Alvarez, and Bravo-Ortega (2012), and Tadesse and Shukralla (2013). In alternative specifications, we also included the real interest rate and its variance to the list. However, these variables never played a role in determining export diversification. Given reduced data availability, we excluded them from the main analysis. Finally, we add the investment share of real GDP per capita (*invest*) to our

5. Government share of GDP is taken from the Penn World Tables 6.3 (PWT). Given the recent debate regarding versions of the PWT, we prefer the 6.3 version and follow Breton (2012). However, our methodology of averaging over many years should alleviate the suggested methodological problem, as put forth by Johnson et al. (2013) in the context of economic growth.

6. We also considered the education variables from Cohen and Soto (2007), but their inclusion would result in the loss of 18 sample countries. Further, pure correlation of their education variables and the respective ones used here is at least 80%.

7. Including total trade as a percentage of GDP, in addition to exports and imports, leads to multicollinearity problems. We also experimented with using total trade (measured as exports plus imports divided by GDP), as opposed to exports and imports separately, but the variable never reached posterior inclusion probabilities close to 50%.

analysis, following Ben Hammouda et al. (2006) and Dennis and Shepherd (2011).

D. Cultural Factors

The uniting theme across cultural variables is that they are mostly fixed over time and largely uncontrollable by policymakers. Various papers include population size as a control variable and dummies for former colonies (Dennis and Shepherd 2011; Parteka and Tamberi 2011). We add the logarithm of total population and dummies for former British, Dutch, French, Portuguese, and Spanish colonies to the list of variables. In addition, we include "language fractionalization" and a binary variable indicating whether a country applies the common law system. Although one might not immediately relate these variables to export diversification, their impact on institutions and growth have been demonstrated, for instance in Acemoglu, Johnson, and Robinson (2001). Thus, because these variables affect other major macroeconomic variables, we also test for an effect on export diversification.

E. Geographical Factors

Our last group of variables focuses on factors that are entirely determined by nature and geography. We follow Bebczuk and Berrettoni (2006) by including continental fixed effects and two binary variables for whether the country is an island or landlocked.

TABLE 2
Summary Statistics

Variable	Variable Name	Mean	(Std. Dev.)	Min.	Max.	N
HHI	HHI	.141	(.178)	.004	.946	105
Political Factors						
Polity IV index	*polity*	4.89	(5.725)	−10	10	105
Government size	*gov*	15.553	(5.378)	5.189	36.789	105
Government effectiveness	*goveff*	.07	(.908)	−1.509	2.223	105
Trade freedom	*tradefreedom*	70.165	(10.578)	35.58	85.150	105
Inflation rate	*infl*	6.565	(4.732)	−.063	23.851	105
Duration secondary education	*edu*1	6.346	(.868)	4	8	105
Duration primary education	*edu*2	5.669	(.893)	3	8	105
Secondary school enrollment, gross %	*edu*3	75.099	(29.883)	9.576	138.277	105
Tertiary school enrollment, gross %	*edu*4	32.787	(25.895)	.495	94.998	105
Primary school enrollment, net %	*edu*5	90.038	(11.928)	43.765	99.972	105
OPEC membership	*opec*	.038	(.192)	0	1	105
Macroeconomic Factors						
Value of exports	*exports*	38.945	(18.724)	6.497	108.193	105
Value of imports	*imports*	44.524	(20.237)	12.205	118.362	105
Natural resource rents in % of GDP	*natres*	8.211	(12.622)	.009	52.454	105
% of total roads paved	*roads*	49.42	(32.415)	3.5	100	105
Gross capital formation as % of GDP	*capital*	23.108	(5.773)	8.789	49.142	105
Foreign direct investment	*fdi*	4.006	(2.999)	.039	15.924	105
GDP per capita	*loggdppc*	7.734	(1.581)	4.854	10.611	105
Terms of trade	*tot*	105.542	(20.853)	28.238	172.641	105
Fuel exports	*fuel*	15.531	(25.515)	.001	97.371	105
Investment share of GDP	*invest*	20.795	(10.576)	3.42	64.483	105
Cultural Factors						
Population	*logpop*	16.215	(1.477)	13.404	20.862	105
Former British colony	*british*	.229	(.422)	0	1	105
Former Dutch colony	*dutch*	.048	(.214)	0	1	105
Former French colony	*french*	.2	(.402)	0	1	105
Former Portuguese colony	*portuguese*	.057	(.233)	0	1	105
Former Spanish colony	*spanish*	.162	(.37)	0	1	105
Language fractionalization	*language*	.369	(.283)	.002	.923	105
English common law system	*commonlaw*	.21	(.409)	0	1	105
Geographical Factors						
Africa	*africa*	.286	(.454)	0	1	105
Asia	*asia*	.21	(.409)	0	1	105
Europe	*europe*	.295	(.458)	0	1	105
North America	*namerica*	.057	(.233)	0	1	105
South America and the Caribbean	*smamerica*	.124	(.331)	0	1	105
Island	*island*	.143	(.352)	0	1	105
Landlocked	*landlocked*	.229	(.422)	0	1	105

Note: All variables are calculated as the mean of country averages from 2001 to 2010.

TABLE 3
Data Sources and Calculations

Variable Name	Source	Calculation
HHI	United Nations	ComTrade data set 6-digit level, see paper for details
Political Factors		
polity	Polity IV	Variable *polity*2 measuring level of democracy, ranging from −10 (totally autocratic) to +10 (total democracy)
gov	PWT 6.3	Government share of real GDP per capita
goveff	Worldwide Governance Indicators (1996–2011)	Estimate of governance effectiveness (ranges from approximately −2.5 (weak) to 2.5 (strong) governance performance)
tradefreedom	Index of Economic Freedom	0–100 (the higher the more freedom)
infl	Global Development Network Growth Database	Average inflation rate
*edu*1	World Bank	Duration of secondary education in years
*edu*2	World Bank	Duration of primary education in years
*edu*3	World Bank	School enrollment secondary, gross %
*edu*4	World Bank	School enrollment tertiary, gross %
*edu*5	World Bank	Total enrollment primary, net %
opec	www.opec.org	Dummy for OPEC membership between 2001–2010
Macroeconomic Factors		
exports	World Bank	Exports of goods and services as % of GDP
imports	World Bank	Imports of goods and services as % of GDP
natres	World Bank	Total natural resources rents in % of GDP
roads	World Bank	Roads, paved (% of total roads)
capital	World Bank	Gross capital formation as % of GDP
fdi	World Bank	Foreign direct investment in % of GDP
loggdppc	World Bank	ln(gdp per capita)
tot	World Bank	Net barter terms of trade index (2000 = 100)
fuel	World Bank	Fuel exports (% of merchandise exports)
invest	PWT 6.3	Investment share of real gdp per capita
Cultural Factors		
logpop	World Bank	ln(total population)
british, *dutch*, *french*, *portuguese*, *spanish*	Own	Five dummies for former colonies
language	Alesina et al. (2003)	Language fractionalization
commonlaw	Own	Dummy for English common law system
Geographical Factors		
africa, *asia*, *europe*, *namerica*, *smamerica*	Own	Five continental dummies for Africa, Asia, Europe, North America, South America, and Caribbean
island, *landlocked*		Two dummies for whether a country is an island/landlocked

F. Descriptive Statistics

Tables 2 and 3 show summary statistics and data sources for each variable in our sample. For every country, we calculate the average HHI from annual data between 2001 and 2010. We then repeat this exercise for each potential explanatory variable. Table 4 shows the average HHI of the 105 countries for which we have data on all explanatory variables. We count 30 African, 22 Asian, 31 European, 6 North American, 3 Oceanic, and 13 South American or Caribbean nations. Countries are sorted by HHI, starting with the most diversified nations. We notice that African and Asian nations are among the most concentrated, in addition to Azerbaijan and Venezuela. European and North American economies generally have a more diversified composition of exports.

Tables 2 and 4 also display the broad variety of countries in our sample. For instance, consider the first political variable, the *polity* index. Although the mean score is approximately 4.9, our sample economies range from entirely autocratic (Saudi Arabia or Swaziland) to totally democratic (various Western nations with +10). Small countries are included, such as Bhutan, Cyprus, and Guyana, in addition to highly populated countries such as India, Indonesia, and the United States. Similarly, the sample contains

TABLE 4
Countries by Average HHI from 2001 to 2010, Ranging from Most to Least Diversified (Sorted by Continent)

Country	HHI	Country	HHI	Country	HHI
Africa		Malaysia	.025	Macedonia, FYR	.027
Morocco	.022	Indonesia	.025	Cyprus	.036
Tunisia	.024	India	.032	Ireland	.038
Madagascar	.045	Bangladesh	.042	Georgia	.042
Uganda	.052	Vietnam	.045	Lithuania	.044
Mauritius	.069	Philippines	.069	Albania	.047
Senegal	.071	Mongolia	.158	Armenia	.081
Kenya	.071	Kyrgyz Republic	.161	Belarus	.085
Namibia	.093	Israel	.188	Norway	.207
Togo	.094	Bhutan	.200	Azerbaijan	.544
Cambodia	.097	Oman	.403		
Egypt, Arab Rep.	.098	Kazakhstan	.406	*North America*	
Gambia, The	.130	Kuwait	.433	United States	.007
Swaziland	.135	Saudi Arabia	.564	Mexico	.025
Ethiopia	.148	Iran	.580	Panama	.036
Lesotho	.162	Yemen, Rep.	.648	Dominican Republic	.038
Niger	.203			Honduras	.064
Malawi	.212	*Europe*		Trinidad and Tobago	.141
Cameroon	.227	Italy	.004		
Central African Republic	.237	Poland	.007	*Oceania*	
Benin	.293	France	.007	New Zealand	.016
Algeria	.303	Spain	.009	Australia	.049
Burundi	.306	Slovenia	.010	Fiji	.083
Mozambique	.318	Denmark	.010		
Ghana	.330	Romania	.011	*South America & Caribbean*	
Mauritania	.385	Sweden	.012	Brazil	.016
Guinea	.417	Ukraine	.013	Argentina	.031
Burkina Faso	.433	Belgium	.013	Guatemala	.033
Mali	.495	United Kingdom	.014	Uruguay	.038
Botswana	.510	Portugal	.015	Nicaragua	.056
Guinea-Bissau	.946	Greece	.015	Peru	.077
		Croatia	.017	Chile	.108
Asia		Switzerland	.017	Guyana	.120
Turkey	.009	Bulgaria	.019	Paraguay	.146
Japan	.013	Hungary	.019	Bolivia	.150
Jordan	.017	Finland	.020	Ecuador	.268
Sri Lanka	.020	Estonia	.021	Jamaica	.270
Pakistan	.021	Latvia	.021	Venezuela	.540
Korea, Rep.	.022	Moldova	.027		

poorer nations, such as Ethiopia and Burundi, and rich nations, such as Japan and Norway.

Altogether, we are testing for 36 explanatory variables: 11 political, 10 macroeconomic, 8 cultural, and 7 geographical factors. The variety of these variables is intended to cover every potential aspect of what might influence a country's level of export diversification. The data set is balanced in the sense that our sample only comprises countries for which we have at least one observation for each of the 36 variables in the time frame 2001–2010. However, the initial annual data are not completely balanced, and Table 5 details the availability of all variables. Some countries, most notably Guinea-Bissau, display little data availability among numerous variables and, most importantly for the dependent variable, the HHI. In our main estimations, we choose to include all potentially available countries, although all derived results hold when excluding countries with few observations for several variables within the 2001–2010 timeframe. Specifically, we considered various cutoffs for excluding countries, for instance only considering economies with more than six observations for the HHI (thus excluding Bhutan, Yemen, Rep., Lesotho, and Guinea-Bissau) and excluding countries that consistently lack information across independent variables (adding Burkina Faso and the Central African Republic to the list, for instance). Finally, we also performed estimations, where we excluded countries that show little data availability for those variables, which eventually turned out to be the most important

TABLE 5
Data Availability Main Sample from 2001 to 2010 (105 Countries)

Variable Name	Availability	Missing Observations[a]
HHI	2001–2010	1: 10 countries. 2: 5 countries. 3: 5 countries. 4: Bhutan, Yemen, Rep. 5: Lesotho. 7: Guinea-Bissau
Political Factors		
polity	2001–2010	None
gov	2001–2010	1: 5 countries. 2: Fiji, Trinidad & Tobago. 3: 4 countries. 4: Burkina Faso. 5: Benin, Niger. 8: Guinea-Bissau.
goveff	2002–2010	None
tradefreedom	2001–2010	1: Central African Republic, Macedonia. 5: Burundi. 8: Bhutan.
infl	2001–2008	1: 5 countries. 2: Central African Republic, Namibia. 7: Guinea.
*edu*1	2001–2010	None
*edu*2	2001–2010	None
*edu*3	2001–2010	1: 33 countries. 2: 11 countries. 3: 6 countries. 4: 4 countries. 5: 6 countries. 6: Central African Republic, Oman. 7: Gambia, Honduras, Sri Lanka. 8: Guinea-Bissau.
*edu*4	2001–2010	1: 27 countries. 2: 8 countries. 3: 7 countries. 4: 7 countries. 5: 8 countries. 6: 4 countries. 7: 5 countries. 8: Dominican Republic, Gambia, Togo. 9: Ecuador, Sri Lanka.
*edu*5	2001–2010	1: 44 countries. 2: 15 countries. 3: 8 countries. 4: 8 countries. 5: Bhutan, Brazil, Oman. 6: 4 countries. 7: 5 countries. 8: Latvia, Uganda. 9: Guinea-Bissau.
opec	2001–2010	None
Macroeconomic Factors		
exports	2001–2010	1: 4 countries. 2: Trinidad & Tobago. 3: Iran, Mali. 4: Burkina Faso, Jamaica. 5: Guyana, Niger. 8: Guinea-Bissau.
imports	2001–2010	1: 4 countries. 2: Trinidad & Tobago. 3: Iran, Mali. 4: Burkina Faso, Jamaica. 5: Guyana, Niger. 8: Guinea-Bissau.
natres	2001–2010	1: 5 countries.
roads	2001–2010	1: 8 countries. 2: 11 countries. 3: 6 countries. 4: 10 countries. 5: 6 countries. 6: 6 countries. 7: 10 countries. 8: 17 countries. 9: 23 countries.
capital	2001–2010	1: Bhutan, Central African Republic, Madagascar. 2: Fiji, Oman, Trinidad & Tobago. 3: Cameroon, Iran, Mali. 4: Burkina Faso, Jamaica. 5: Niger. 8: Guinea-Bissau.
fdi	2001–2010	1: Belgium, Iran, Namibia. 2: Bhutan, Gambia.
loggdppc	2001–2010	1: Iran.
tot	2005–2010	4: Cyprus.
fuel	2001–2010	1: 8 countries. 2: 7 countries. 3: 8 countries. 4: Benin, Bhutan, Gambia. 5: Lesotho. 7: Mauritania. 9: Guinea-Bissau.
invest	2001–2007	None
Cultural Factors		
logpop	2001–2010	None
british, *dutch*, *french*, *portuguese*, *spanish*	2001–2010	None
language	2001	None
commonlaw	2001–2010	None
Geographical Factors		
africa, *asia*, *europe*, *namerica*, *smamerica*	2001–2010	None
island, *landlocked*	2001–2010	None

Note: For example, “7: Guinea-Bissau” means that 7 annual observations for Guinea-Bissau are missing.
[a]Listing the number of annual observations that are missing for a country.

ones. However, all derived results remain robust to these extensions (available upon request).

IV. RESULTS

Following our methodological outline given in Section II, we apply BMA by using the birth/death algorithm.[8] The number of iteration draws to be sampled (ex burn-ins) becomes 1,200,000 and the number of burn-in draws for

8. All estimations are performed with the R package (R Development Core Team 2011), using the BMS library (Feldkircher and Zeugner 2012) and IVBMA library (Lenkoski, Karl, and Neudecker 2013).

the MCMC sampler is 2,000,000. With the algorithm visiting 28,068 of 2^{36} potential models, the correlation between the analytical and MCMC posterior model probabilities turns out to be above 1.00, indicating a good performance of the algorithm (Fernandez, Ley, and Steel 2001a).

A. *Main BMA Results*

Table 6 shows our main BMA results. By far, the two most important predictors of export diversification are total net enrollment in primary education (*edu*5) and total natural resource rents as a percentage of GDP (*natres*). The PIP of both variables is outstanding with values of .997 and .986, meaning that they are included in 99.7% and 98.6% of the best models. After that, we notice a remarkable drop in the PIP toward less than 44% for the binary variable indicating Portuguese colonies. Moving to the fourth most important variable, we then see another steep drop to PIP's under .03%.

Relying on the common threshold levels provided by the respective literature then provides us with a clear answer (Eicher, Henn, and Papageorgiou 2012; Kass and Raftery 1995): there exists decisive evidence for the importance of *edu*5 and natres (PIP > 98%), but evidence against all other 34 factors (PIP < 50%). The upper graph in Figure 1 displays the posterior model size distribution and provides further evidence that the level of export diversification is mostly determined by these two variables. Table 7 confirms that the PMP of using only these two variables comes out to be over 46%. The second-best model then adds a dummy for Portuguese colonies and reaches a PMP of approximately 34%. After that, model probabilities decline substantially to approximately 1%.

In addition, column 5 (Cond. Pos. Sign) of Table 6 allows us to conclude the sign of the respective coefficients. For instance, a value of 1 means that a variable is likely to have a positive effect on a country's HHI in all models and thus should decrease export diversification. Primary enrollment rates, on the other hand, are suggested to increase diversification levels, as indicated by the probability of getting a positive sign being equal to 0. Looking at *natres*, we find conclusive evidence that countries in which natural resources play a strong role find it harder to diversify their exports (Cond. Pos. Sign of 1).

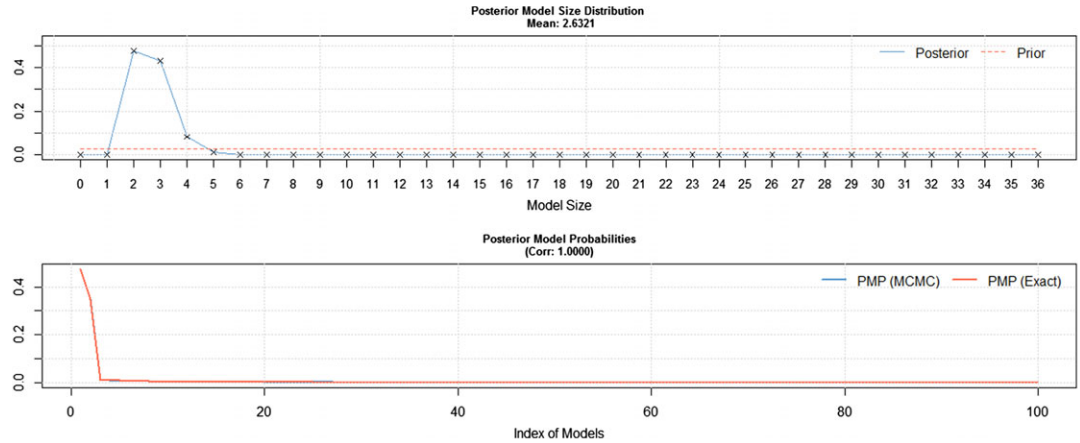
Finally, Figure 2 displays the posterior probability density functions of the coefficients associated with the two dominant variables, given that the corresponding variable is included in the regression. For instance, the median and the expected value of the density function of the coefficient associated with natural resource abundance in GDP is approximately .0086, while the two vertical dashed lines show a two times conditional posterior standard deviation interval. Figure 2 shows that the marginal effect of natural resources in GDP has a positive relationship with the HHI, thus implying less diversification. The density function of primary enrollment, however, is firmly negative, with a suggested coefficient of −.0061: higher primary enrollment rates are suggested to lower the HHI and therefore encourage diversification.

The importance of natural resources in the context of export diversification is no surprise and

TABLE 6
BMA: Main Results Averaging All Variables from 2001 to 2010 (105 Countries)

Variable	PIP	Post Mean	Post SD	Cond. Pos. Sign
edu5	.9967	−.0061	.0011	.0000
natres	.9859	.0086	.0014	1.0000
portuguese	.4320	.0693	.0852	1.0000
logpop	.0262	−.0004	.0028	.0000
edu1	.0258	−.0007	.0047	.0000
fuel	.0229	.0001	.0005	1.0000
africa	.0176	.0013	.0112	1.0000
language	.0104	.0007	.0081	1.0000
namerica	.0080	−.0006	.0077	.0000
french	.0070	−.0003	.0043	.0000
infl	.0064	.0000	.0003	1.0000
goveff	.0060	−.0001	.0018	.0000
edu3	.0056	.0000	.0001	.0279
island	.0049	−.0001	.0030	.0000
edu2	.0049	.0001	.0013	.9919
loggdpcap	.0048	.0000	.0010	.0079
dutch	.0047	−.0002	.0044	.0000
imports	.0041	.0000	.0000	.9298
landlocked	.0040	.0001	.0024	1.0000
spanish	.0040	.0000	.0022	.0368
europe	.0040	−.0001	.0023	.0273
opec	.0039	.0003	.0060	1.0000
fdi	.0038	.0000	.0003	.0051
exports	.0037	.0000	.0001	.0538
capital	.0036	.0000	.0002	.0000
gov	.0035	.0000	.0002	.9914
polity	.0034	.0000	.0001	.9165
british	.0033	.0000	.0016	.4322
edu4	.0030	.0000	.0000	.0362
tradefreedom	.0030	.0000	.0001	.0820
tot	.0030	.0000	.0000	.0117
invest	.0029	.0000	.0001	.4158
commonlaw	.0029	.0000	.0015	.6110
roads	.0028	.0000	.0000	.0672
smamerica	.0027	.0000	.0019	.4619
asia	.0025	.0000	.0015	.0806

FIGURE 1
BMA: Prior and Posterior Model Size



confirms numerous previous analyses of natural resource dependence hampering export diversification (see Bebczuk and Berrettoni 2006; Sachs and Warner 2001; van der Ploeg 2011, for a recent discussion). In terms of education, it is interesting to see the importance of the enrollment percentage (not duration) of primary education (not secondary or tertiary) to be highly important. Thus, a broad base of basic education seems to matter for export diversification in the long run. This result confirms the general importance of schooling, as found in Agosin, Alvarez, and Bravo-Ortega (2012), who control for average years of schooling. As we are including five different aspects of education, we are able to be more specific and conclude that it is the primary enrollment rate that leads to export diversification. We also wish to note that these BMA results are robust to choosing different timeframes or cutoff points. For instance, we derive the same qualitative results when excluding the time period after the global financial crisis in 2007. In all these estimations, the remaining variables return a PIP of under .5 (results available upon request).[9]

9. We also experimented with using other data sources, such as the Feenstra database (Feenstra et al. 2005), but ran into substantial problems with data availability of trade (the Feenstra database is available until the year 2000) and a variety of explanatory variables. This concerns numerous variables such as trade freedom, government effectiveness, and roads, but also the variables, which turned out to be important in describing export diversification (primary enrollment rates and the importance of natural resources). In general, the further we go back in time, the less data becomes available.

Finally, we want to mention that we also estimated various alternative specifications, such as using lagged values from 1991 to 2000 for all explanatory variables or dropping countries with less annual observations for trade.

B. *Notable Country Examples*

We now turn to a brief overview of some countries with noteworthy levels of export diversification and their values for the most important factors. Table 8 displays the ten most concentrated and the ten most diversified export baskets in our sample. In addition, we display sample means and medians of the most important variables at the bottom of the table. We notice that six out of the ten least diversified export baskets belong to countries with a value of *natres* over 30%, which is remarkable compared to the sample median of 2.5 and a mean of 8.2%. In addition, four of these countries show primary enrollment rates of 75% or less, compared to a sample mean of 90%. Thus, for these most concentrated countries, the main reasons can be found within the two most important variables. The only outlier in this group is Botswana with a *natres* value of 3.4% and an average primary enrollment rate of over 84.6%. The difference here might be that the World Bank does not include diamonds in their definition of natural resources, and Botswana functions as a major exporter of diamonds on the world market.

Turning to the most diversified nations, resources only play a very minor role for most

TABLE 7
BMA: Variables Included in 10 Top Models Averaging All Variables from 2001 to 2010 (105 Countries)

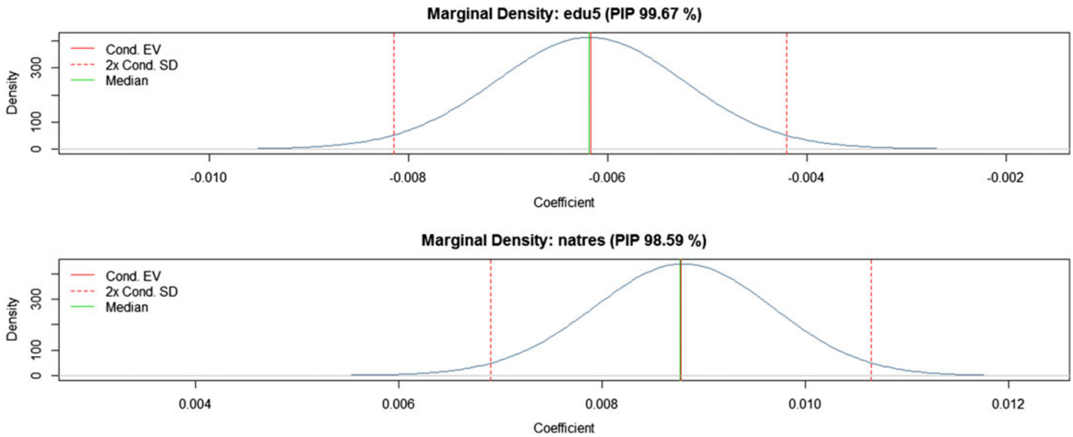
Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
edu5	1	1	1	1	1	1	1	1	1	1
natres	1	1	1	1	1	1	1	0	0	1
portuguese	0	1	0	1	0	1	0	0	1	0
edu1	0	0	1	0	0	1	0	0	0	0
logpop	0	0	0	1	0	0	1	0	0	0
africa	0	0	0	0	1	0	0	0	0	0
fuel	0	0	0	0	0	0	0	1	1	0
language	0	0	0	0	0	0	0	0	0	1
island	0	0	0	0	0	0	0	0	0	0
landlocked	0	0	0	0	0	0	0	0	0	0
opec	0	0	0	0	0	0	0	0	0	0
british	0	0	0	0	0	0	0	0	0	0
dutch	0	0	0	0	0	0	0	0	0	0
french	0	0	0	0	0	0	0	0	0	0
spanish	0	0	0	0	0	0	0	0	0	0
commonlaw	0	0	0	0	0	0	0	0	0	0
europe	0	0	0	0	0	0	0	0	0	0
asia	0	0	0	0	0	0	0	0	0	0
namerica	0	0	0	0	0	0	0	0	0	0
smamerica	0	0	0	0	0	0	0	0	0	0
polity	0	0	0	0	0	0	0	0	0	0
gov	0	0	0	0	0	0	0	0	0	0
goveff	0	0	0	0	0	0	0	0	0	0
tradefreedom	0	0	0	0	0	0	0	0	0	0
infl	0	0	0	0	0	0	0	0	0	0
edu2	0	0	0	0	0	0	0	0	0	0
edu3	0	0	0	0	0	0	0	0	0	0
edu4	0	0	0	0	0	0	0	0	0	0
exports	0	0	0	0	0	0	0	0	0	0
imports	0	0	0	0	0	0	0	0	0	0
capital	0	0	0	0	0	0	0	0	0	0
fdi	0	0	0	0	0	0	0	0	0	0
loggdpcap	0	0	0	0	0	0	0	0	0	0
tot	0	0	0	0	0	0	0	0	0	0
invest	0	0	0	0	0	0	0	0	0	0
roads	0	0	0	0	0	0	0	0	0	0
PMP (MCMC)	.4657	.3432	.0119	.0117	.0088	.0077	.0080	.0058	.0048	.0047
PMP (Exact)	.4643	.3409	.0120	.0115	.0093	.0077	.0074	.0070	.0053	.0044

of them, corresponding to less than 1% of GDP. In terms of primary education, all of the ten most diversified countries show enrollment rates of greater than 95%. Although these are snapshots of the data, we notice a general pattern. In fact, the pure correlation between country-level averages of the HHI with *natres* is greater than 63%, whereas the correlation with *edu*5 stands at −45% (not displayed). However, the correlation between *edu*5 and *natres* lies under −5%, suggesting that resource-abundant nations do not necessarily have low primary enrollment rates.

C. Excluding OECD Countries

Looking at Table 8, we can note another pattern when comparing the most diversified export baskets to the most concentrated ones: with the exception of Romania, nine out of the ten most diversified nations are OECD countries, whereas none of the least diversified nations are members of the OECD. In fact, if we showed an additional column of GDP per capita, we would also note a substantial difference in income levels between those two groups of countries. Thus, one concern regarding our results could be that unobserved factors are driving our findings, or that OECD nations are simply more diversified for reasons outside our potential explanatory variables. Addressing this concern, Table 9 shows BMA results when excluding all OECD countries from the sample. Thus, if any results were driven by OECD nations, we would not observe

FIGURE 2
BMA: Posterior Probability Density of the Most Important Coefficients



the strong results of our initial estimations. However, we quickly notice that the findings receive strong confirmation, as both *edu*5 and *natres* receive PIPs over .93 and *portuguese* remains in a distant third place, this time with a PIP of .17. Thus, OECD nations are not driving our results.

D. Addressing Endogeneity

Another concern in the literature surrounding the determinants of export diversification centers around potential reverse causality problems. This issue concerns especially income per capita, as export diversification may well have an independent effect on a country's income levels (see Agosin 2009; Cadot, Carrère, and Strauss-Kahn 2012; Lederman and Maloney 2003). However, if we also assume that richer countries may choose to diversify their exports, potentially to avoid being subject to volatility in specific markets, then a common regression framework may suffer from reverse causality issues. Thus, endogeneity might be present between GDP per capita (*loggdpcap*) and the HHI. In order to address this problem, we apply an IVBMA framework. In this case, the first stage uses *loggdpcap* as the dependent variable and considers various geographical, cultural, and colonial factors as instrumental variables. In particular, we employ the variables *island*, *landlocked*, *europe*, *asia*, *africa*, *namerica*, *smamerica*, *language*, *commonlaw*, *opec*, *british*, *dutch*, *french*, *spanish*, and *portuguese*. We select those variables as instruments because they are mostly impossible to be influenced by income levels and their effects on income have been shown previously (for instance consider the recent evidence on the relationship between colonial heritage and income levels, initiated by Acemoglu, Johnson, and Robinson 2001).

Table 10 displays the first- and second-stage results from using the IVBMA mechanism. As before, the main results show that natural resources and primary enrollment rates are suggested to be the only significant factors determining export diversification. In this case, *natres* returns a PIP of .96, whereas *edu*5 only barely loses ground toward a PIP of .80. We note that using an instrumental variable approach marginally increases the importance of GDP per capita (PIP of .15), but the variable remains far from being significant if we assume a minimal threshold level of PIP = .50.

We wish to note one shortcoming of our BMA and IVBMA approaches regarding the sample structure. Although our analysis is able to control for numerous potential determinants, we are unable to control for unobservable country-specific aspects, due to our long-term outline of averaging variables over a decade. In fact, we also experimented with a panel BMA approach (following Moral-Benito 2013a), using annual data for all variables. However, data availability leads to a substantially reduced number of observations for numerous countries. In order to use annual data, one needs the information for *all* variables in a specific country-year observation and, as Table 5 shows, there are some variables, which only display information for certain years

TABLE 8
The 10 Least and the 10 Most Diversified Export Baskets

Country	Average HHI 2001–2010	Average Natural Resource Rents (% of GDP) 2001–2010	Average Primary Enrollment Rate 2001–2010
10 Least Diversified Export Baskets			
Guinea-Bissau	.946	6.6	75.0
Yemen, Rep.	.648	31.1	73.9
Iran	.580	42.0	96.6
Saudi Arabia	.564	51.2	85.2
Azerbaijan	.544	52.5	84.4
Venezuela	.540	32.6	93.9
Botswana	.510	3.4	84.6
Mali	.495	7.7	55.9
Kuwait	.433	51.9	98.1
Burkina Faso	.433	4.1	45.7
10 Most Diversified Export Baskets			
Italy	.004	.2	99.4
Poland	.007	.9	96.3
United States	.007	1.2	96.2
France	.007	.1	99.4
Turkey	.009	.4	97.1
Spain	.009	.1	99.8
Slovenia	.010	.2	96.7
Denmark	.010	2.5	97.4
Romania	.011	3.8	95.7
Sweden	.012	.8	97.8
Mean	.141	8.2	90.0
Median	.056	2.5	94.8

between 2001 and 2010 (for instance *goveff* with 2002–2010 or *invest* with 2001–2007). This naturally decreases the available data below the ideal sample of ten observations per country, and eventually many countries end up with less than four yearly observations in this exercise.

V. CONCLUSIONS

This paper seeks to provide a thorough understanding of the true determinants of export diversification levels. With the lack of a comprehensive theoretical foundation and a variety of potential factors suggested in the empirical literature, finding the true factors associated with export diversification becomes difficult. However, using BMA allows us to get to the grain of model uncertainty. In addition, a recently developed instrumental variable BMA method (IVBMA) allows us to control for potential reverse causality between income and export diversification levels.

Overall, our sample considers 36 potential determinants of export diversification and consists of 105 countries. This means that there are 68,719,476,736 (2^{36}) variable combinations under which the true model is hiding. The results from BMA estimations are remarkably strong in suggesting only two variables to play a decisive role in predicting export diversification levels: a higher share of primary enrollment rates encourages export diversification (99.7% PIP in the true model), whereas the share of natural resources in GDP lowers export diversification (98.6%). The importance of natural resources in this context comes as no surprise, but the role of primary education coverage does. Although human capital in general has been mentioned (e.g., Agosin, Alvarez, and Bravo-Ortega 2012), our findings highlight the role of primary enrollment rates. Other education variables, such as secondary and tertiary schooling, or the duration of these, do not seem to matter. Even more surprisingly, the development level (GDP per capita), as well as political variables (e.g., regime form, freedom to trade, or government effectiveness) never

TABLE 9
BMA: Results from Averaging All Variables from 2001 to 2010, Only Using Non-OECD Countries (79 Countries)

Variable	PIP	Post Mean	Post SD	Cond. Pos. Sign.
edu5	.9623	−5.9760E-03	1.6998E-03	.0000
natres	.9319	8.1301E-03	2.4534E-03	1.0000
portuguese	.1685	2.7398E-02	6.4791E-02	1.0000
fuel	.0725	2.8945E-04	1.0654E-03	.9999
africa	.0230	2.3162E-03	1.6912E-02	1.0000
edu1	.0196	−6.4377E-04	5.0952E-03	.0000
logpop	.0174	−3.5740E-04	3.0440E-03	.0000
edu3	.0171	−3.5096E-05	2.9579E-04	.0038
goveff	.0148	−8.1713E-04	7.6406E-03	.0000
loggdpcap	.0140	−5.5597E-04	5.3761E-03	.0000
language	.0070	5.2497E-04	7.9041E-03	1.0000
namerica	.0066	−5.9529E-04	9.1629E-03	.0000
edu2	.0059	1.0878E-04	1.9498E-03	.9830
infl	.0057	2.2923E-05	3.9201E-04	1.0000
french	.0056	−2.2616E-04	4.0926E-03	.0001
roads	.0048	−2.5332E-06	5.5204E-05	.0045
island	.0046	−1.9294E-04	4.2109E-03	.0000
europe	.0043	−1.6246E-04	3.7446E-03	.0154
asia	.0043	−1.0696E-04	3.0662E-03	.0251
fdi	.0040	−1.7790E-05	4.6072E-04	.0412
invest	.0039	−2.3372E-06	1.0877E-04	.0629
landlocked	.0038	8.2382E-05	2.6866E-03	.9872
dutch	.0035	−1.5130E-04	4.4882E-03	.0021
exports	.0035	−2.2168E-06	6.9455E-05	.0218
capital	.0034	−5.8227E-06	1.7662E-04	.0012
british	.0033	−3.4902E-05	2.2899E-03	.0651
edu4	.0032	−2.7710E-06	8.5607E-05	.0066
tradefreedom	.0031	−2.2725E-06	1.0317E-04	.0488
smamerica	.0031	6.2168E-05	2.6778E-03	.9869
commonlaw	.0030	−2.0457E-05	2.2561E-03	.2356
imports	.0030	7.9687E-07	4.9590E-05	.9452
spanish	.0030	−1.7874E-05	2.3065E-03	.0617
gov	.0027	2.9200E-06	1.6451E-04	.9838
opec	.0026	1.6434E-04	5.3744E-03	.9860
polity	.0024	1.4015E-06	1.5152E-04	.9440
tot	.0022	−1.9278E-07	4.6820E-05	.1263

TABLE 10
IVBMA: Results Averaging All Variables from 2001 to 2010 (105 Countries)

Second Stage Results			First Stage Results Estimating *loggdpcap*		
Variable	**PIP**	**Coefficient**	**Variable**	**PIP**	**Coefficient**
natres	.9571	.0083	*goveff*	1.0000	.6533
edu5	.8018	−.0050	*edu2*	1.0000	.5184
loggdpcap	.1533	−.0057	*edu3*	1.0000	.0287
edu2	.0927	.0031	*edu1*	.9996	.3465
goveff	.0785	−.0028	*dutch*	.9430	−.6875
edu1	.0556	−.0015	*namerica*	.8193	.5724
fuel	.0453	.0002	*opec*	.7220	.4765
logpop	.0265	−.0004	*language*	.6481	−.2998
edu3	.0090	.0000	*commonlaw*	.4945	−.1479
fdi	.0079	.0000	*smamerica*	.4744	.1985
infl	.0068	.0000	*spanish*	.3864	.1205
gov	.0066	.0000	*africa*	.3853	−.0969
polity	.0054	.0000	*landlocked*	.3651	−.0850
capital	.0047	.0000	*europe*	.2765	.0729
tradefreedom	.0029	.0000	*asia*	.2690	.0731
invest	.0025	.0000	*british*	.2095	−.0178
edu4	.0014	.0000	*portuguese*	.2007	−.0042
imports	.0004	.0000	*fdi*	.1968	−.0083
exports	.0002	.0000	*imports*	.1780	−.0032
tot	.0002	.0000	*island*	.1632	−.0017
roads	.0001	.0000	*french*	.1551	.0043
			exports	.1399	.0026
			polity	.1071	−.0032
			tradefreedom	.0721	.0012
			logpop	.0654	.0026
			natres	.0416	.0005
			fuel	.0407	.0003
			capital	.0365	−.0007
			gov	.0253	.0001
			roads	.0241	.0001
			infl	.0229	.0002
			invest	.0160	.0001
			edu5	.0115	.0001
			edu4	.0076	.0000
			tot	.0044	.0000

receive support throughout this study. Taking into account the potential reverse causality with export diversification—not only may richer countries diversify more, but more diversified countries may also be richer—in an IVBMA framework confirms our results. Thus, our findings suggest that the strong relationship between income levels and export diversification could be spurious and especially driven by the presence of natural resources and basic primary enrollment rates. Similarly, geographical aspects, such as being landlocked or an island country, do not seem to matter for export diversification.

Our results underline the importance of broad basic education levels in the context of export diversification—a factor that has previously been pointed out in the context of economic growth, for instance by Psacharopoulos (1994) or Sala-i Martin, Doppelhofer, and Miller (2004). We believe that these results are an important step toward understanding the true roots of export diversification over the long run. If, as previous works suggest, a diversified export basket indeed lowers income volatility, generates knowledge spillovers, and opens doors to a smoother development path, then our paper provides support for the importance of primary education coverage to achieve these goals.

REFERENCES

Acemoglu, D., S. Johnson, and J. A. Robinson. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5), 2001, 1369–401.

Agosin, M. *Trade and Growth: Why Asia Grows Faster than Latin America*. Washington, DC: IDB Publications, 2006.

———. "Export Diversification and Growth in Emerging Economies." *CEPAL Review*, 97, 2009, 115–31.

Agosin, M. R., R. Alvarez, and C. Bravo-Ortega. "Determinants of Export Diversification around the World:

1962–2000." *The World Economy*, 35(3), 2012, 295–315.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg. "Fractionalization." *Journal of Economic Growth*, 8(2), 2003, 155–94.

Al-Marhubi, F. "Export Diversification and Growth: An Empirical Investigation." *Applied Economics Letters*, 7(9), 2000, 559–62.

Alwang, J., and P. Siegel. "Portfolio Models and Planning for Export Diversification: Malawi, Tanzania and Zimbabwe." *Journal of Development Studies*, 30(2), 1994, 405–22.

Amin Gutiérrez de Piñeres, S., and M. Ferrantino. "Export Diversification and Structural Dynamics in the Growth Process: The Case of Chile." *Journal of Development Economics*, 52(2), 1997, 375–91.

Amurgo-Pacheco, A. *Patterns of Export Diversification in Developing Countries: Intensive and Extensive Margins*, Vol. 4473. Washington, DC: World Bank Publications, 2008.

Barro, R., and J. Lee. "International Data on Educational Attainment Updates and Implications." NBER Working Paper Series 7911, 2000.

Bebczuk, R., and N. Berrettoni. "Explaining Export Diversification: An Empirical Analysis." Documento de Trabajo Nro 65. 2006.

Beine, M., and S. Coulombe. "Economic Integration and the Diversification of Regional Exports: Evidence from the Canadian–US Free Trade Agreement." *Journal of Economic Geography*, 7(1), 2007, 93–111.

Ben Hammouda, H., S. Karingi, A. Njuguna, and M. Sadni Jallab. "Diversification: Towards a New Paradigm for Africa's Development." ATPC Work in Progress No. 35. 2006.

Breton, T. R. "Penn World Table 7.0: Are the Data Flawed?" *Economics Letters*, 117(1), 2012, 208–10.

Cadot, O., C. Carrère, and V. Strauss-Kahn. "Export Diversification: What's Behind the Hump?" *Review of Economics and Statistics*, 93(2), 2011, 590–605.

———."Trade Diversification, Income, and Growth: What Do We Know?" *Journal of Economic Surveys*, 27(4), 2012, 790–812.

Chaney, T. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *American Economic Review*, 98(4), 2008, 1707–21.

Cohen, D., and M. Soto. "Growth and Human Capital: Good Data, Good Results." *Journal of Economic Growth*, 12(1), 2007, 51–76.

Dennis, A., and B. Shepherd. "Trade Facilitation and Export Diversification." *The World Economy*, 34(1), 2011, 101–22.

Eicher, T. S., L. Helfman, and A. Lenkoski. "Robust FDI Determinants: Bayesian Model Averaging in the Presence of Selection Bias." *Journal of Macroeconomics*, 34(3), 2012a, 637–51.

Eicher, T. S., C. Henn, and C. Papageorgiou. "Trade Creation and Diversion Revisited: Accounting for Model Uncertainty and Natural Trading Partner Effects." *Journal of Applied Econometrics*, 27(2), 2012b, 296–321.

Feenstra, R. C., R. E. Lipsey, H. Deng, A. C. Ma, and H. Mo. "World Trade Flows: 1962–2000." Technical Report, National Bureau of Economic Research, 2005.

Feldkircher, M., and S. Zeugner. *Bayesian Model Averaging*. Vienna, Austria: R Foundation for Statistical Computing, 2012.

Fernandez, C., E. Ley, and M. Steel. "Benchmark Priors for Bayesian Model Averaging." *Journal of Econometrics*, 100, 2001, 381–427.

———."Model Uncertainty in Cross-Country Growth Regressions." *Journal of Applied Econometrics*, 16(5), 2001, 563–76.

Gylfason, T. "Development and Growth in Mineral-Rich countries." *Sustainable Growth and Resource Productivity: Economic and Global Policy Issues*, 1(88), 2009, 42–84.

Hastings, W. "Monte Carlo Sampling Methods Using Markov Chains and Their Application." *Biometrika*, 57, 1970, 97–109.

Hausmann, R., J. Hwang, and D. Rodrik. "What You Export Matters." *Journal of Economic Growth*, 12(1), 2007, 1–25.

Herzer, D., and F. D. Nowak-Lehnmann. "What Does Export Diversification Do for Growth? An Econometric Analysis." *Applied Economics*, 38(15), 2006, 1825–38.

Imbs, J., and R. Wacziarg. "Stages of Diversification." *American Economic Review*, 93(1), 2003, 63–86.

Jansen, M. "Income Volatility in Small and Developing Economies: Export Concentration Matters." Discussion Paper, World Trade Organization No. 3 2004.

Johnson, S., W. Larson, C. Papageorgiou, and A. Subramanian. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates." *Journal of Monetary Economics*, 60(2), 2013, 255–74.

Karl, A., and A. Lenkoski. "Instrumental Variable Bayesian Model Averaging via Conditional Bayes Factor." Technical Report, Heidelberg University, 2012.

Kass, R. E., and A. E. Raftery. "Bayes Factors." *Journal of the American Statistical Association*, 90(430), 1995, 773–95.

Koop, G., D. J. Poirier, and J. L. Tobias. *Bayesian Econometric Methods*, Vol. 7. Cambridge: Cambridge University Press, 2007.

Koop, G., R. León-Gonzalez, and R. Strachan. "Bayesian Model Averaging in the Instrumental Variable Regression Model." *Journal of Econometrics*, 171, 2012, 237–50.

Leamer, E. E. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley & Sons, 1978.

Lederman, D., and W. Maloney. "Trade Structure and Growth." World Bank Policy Research Working Paper No. 3025, 2003.

Lenkoski, A., A. Karl, and A. Neudecker. *Bayesian Instrumental Variable Estimation and Model Determination via Conditional Bayes Factors*. Vienna, Austria: R Foundation for Statistical Computing, 2013.

Ley, E., and M. Steel. "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions." *Journal of Applied Econometrics*, 24(4), 2009, 651–74.

Madigan, D., and A. Raftery. "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window." *Journal of the American Statistical Association*, 89, 1994, 1535–46.

Madigan, D., J. York, and D. Allard. "Bayesian Graphical Models for Discrete Data." *International Statistical Review*, 63(2), 1995, 215–32.

Mengistu, A. "Do Physical and Human Capital Matter for Export Diversification? A Comparative Analysis of Sub-Saharan Africa and East Asia." *African and Asian Studies*, 8(1–2), 2009, 1–2.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. "Equations of State Calculations by Fast Computing Machines." *The Journal of Chemical Physics*, 21, 1953, 1087–92.

Moral-Benito, E. "Determinants of Economic Growth: A Bayesian Panel Data Approach." *Review of Economics and Statistics*, 94(2), 2013, 566–79.

———."Model Averaging in Economics: An Overview." *Journal of Economic Surveys*, 29, 2013, 46–75.

Newfarmer, R., W. Shaw, and P. Walkenhorst. *Breaking into New Markets: Emerging Lessons for Export*

Diversification. Washington, DC: World Bank Publications, 2009.

Parteka, A. "Employment and Export Specialisation along the Development Path: Some Robust Evidence." *Review of World Economics*, 145(4), 2010, 615–40.

Parteka, A., and M. Tamberi. "Determinants of Export Diversification: An Empirical Investigation." Working Paper No. 327, Universita Politecnica delle Marche, Dipartimento di Economia, 2008.

———."Export Diversification and Development-Empirical Assessment." Working Paper No. 359, Università Politecnica delle Marche, Dipartimento di Scienze Economiche e Sociali, 2011.

Petersson, L. "Export Diversification and Intra-Industry Trade in South Africa." *South African Journal of Economics*, 73(4), 2005, 785–802.

van der Ploeg, F. "Natural Resources: Curse or Blessing?" *Journal of Economic Literature*, 49(2), 2011, 366–420.

Psacharopoulos, G. "Returns to Investment in Education: A Global Update." *World Development*, 22(9), 1994, 1325–43.

Raftery, A. "Bayesian Model Selection in Social Research." *Sociological Methodology*, 25, 1995, 111–63.

Raftery, A., and Y. Zheng. "Discussion: Performance of Bayesian Model Averaging." *Journal of the American Statistical Association*, 98(464), 2003, 931–98.

Raftery, A., D. Madigan, and J. Hoeting. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association*, 92(437), 1997, 179–91.

R Development Core Team. *R: A Language and Environment for Statistical Computing. R. Version 2.13.0*. Vienna, Austria: R Foundation for Statistical Computing, 2011.

Sachs, J., and A. Warner. "The Curse of Natural Resources." *European Economic Review*, 45(4), 2001, 827–38.

Sala-i Martin, X., G. Doppelhofer, and R. I. Miller. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review*, 94, 2004, 813–835.

Tadesse, B., and E. Shukralla. "The Impact of Foreign Direct Investment on Horizontal Export Diversification: Empirical Evidence." *Applied Economics*, 45(2), 2013, 141–59.

Taylor, T. "Export Diversification in Latin America and the Caribbean." *The International Trade Journal*, 17(2), 2003, 101–28.

Volpe Martincus, C., and J. Carballo. "Is Export Promotion Effective in Developing Countries? Firm-Level Evidence on the Intensive and the Extensive Margins of Exports." *Journal of International Economics*, 76(1), 2008, 89–106.