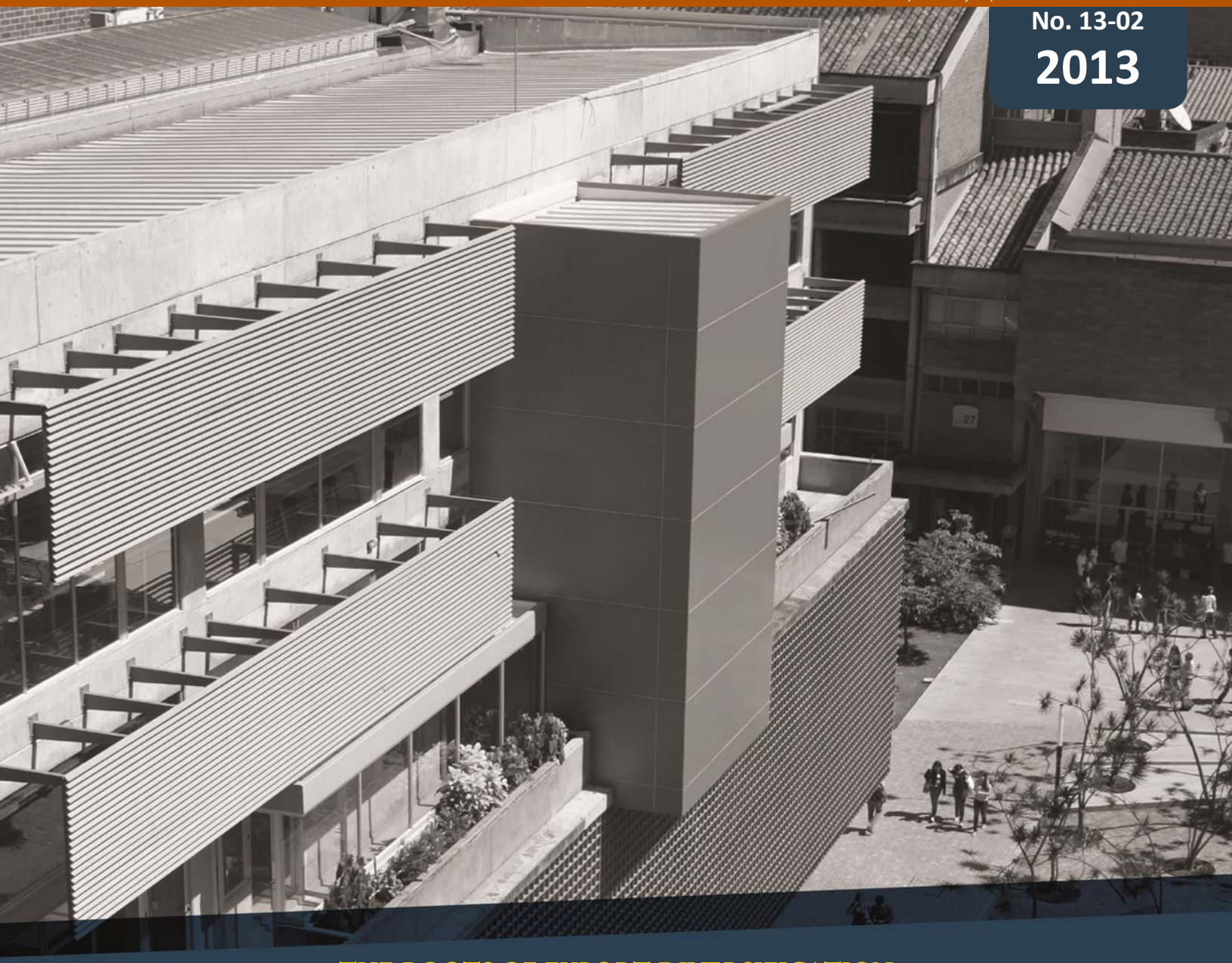
No. 13-02

2013



THE ROOTS OF EXPORT DIVERSIFICATION

Michael Jetter

Andrés Ramírez Hassan

Documentos de trabajo

Economía y Finanzas

Centro de Investigaciones Económicas y Financieras (CIEF)

UNIVERSIDAD EAFIT

Abierta al mundo

Foto: Juan Carlos Ramírez

The Roots of Export Diversification*

Michael Jetter† Andrés Ramírez Hassan‡

January 22, 2013

Abstract

Countries with diversified export baskets take advantage of various benefits, which are said to foster and stabilize economic growth directly and through indirect channels (e.g. reduced income volatility, positive externalities, spillover effects). This is especially important in the context of developing economies. However, identifying the true determinants of export diversification is difficult as there exists no comprehensive theoretical or empirical framework to capture all potential factors in their entirety. This paper uses Bayesian Model Averaging to uncover the true long-term roots of export diversification among 43 potential determinants, and thus 2^{43} potential models. Our results suggest that only four factors are important in predicting export diversification levels over the long run: natural resource rents as a percentage of GDP (100 % posterior inclusion probability), primary school enrollment rates (96 %), population size (25 %), and foreign direct investment levels (17 %). Many prominent candidates turn out to be insignificant in determining diversification levels. Neither policy-related variables (e.g. tariffs, freedom from trade regulations or democracy) nor macroeconomic factors (such as trade openness, terms of trade or domestic investment levels) nor geographical remoteness (whether the country is an island or landlocked) play a role. Various robustness checks confirm our results.

JEL Classification: C11, F1, O11
Keywords: Export Diversification, International Trade, Bayesian Model Averaging

*We thank Theodore R. Breton, Gustavo J. Canavire Bacarreza, Matthew T. Cole, Volodymyr Lugovskyy, and Jay Walker for fruitful comments and discussions. Also, we are grateful for excellent research assistance from Elizabeth Montoya Betancur. All remaining errors are our own.

†Department of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia; email: mjetter7@gmail.com; web: www.michaeljetter.com

‡Departament of Economics, School of Economics and Finance, Universidad EAFIT, Medellín, Colombia. email: aramir21@eafit.edu.co

1 Introduction

As globalization has taken unforeseen dimensions over the past decades, various aspects of international trade are gaining importance. Economists and politicians not only discuss trade openness anymore, but also the form and diversity of a country's export basket. Influential works by Imbs and Wacziarg (2003) and Cadot et al. (2011) describe the pattern of export diversification over development levels of countries. A diversified export basket is said to foster economic growth in several ways: (*i*) reduced income volatility in the exporting sector by less exposure to international shocks in particular industries, (*ii*) externalities from learning-by-doing or learning-by-exporting, and (*iii*) potential spillover effects.[1]

Although the effects from export diversification are receiving substantial attention, surprisingly little research has investigated its roots. But if diversifying exports carries benefits in the development of an economy, then it is useful to know what determines diversification levels over the long run. Simply comparing common diversification indices between countries is tempting, yet might result in misleading conclusions if one ignores an economy's unique capacities to diversify. For example, if the real determinants lie in the range of policies (e.g. trade regulations or education levels), then the degree of diversification can be influenced. On the other hand, if country-specific conditions (such as geographical location, population size, or being a former colony) play a dominant role, little remains to be done for policymakers.

The novelty of our approach lies in the application of Bayesian Model Averaging (BMA) to a long list of potential export diversification determinants, in order to address the underlying problem of model uncertainty. The reasons for our approach are threefold: (1) Given the complexity of the topic, there exists no theoretical foundation suggesting a comprehensive list of export diversification determinants. (2) Previous empirical research could not agree on a unique set of determinants and control variables. (3) The presence of short-term fluctuations and po-

[1]For the volatility argument, see Agosin (2006), Agosin (2009), or Jansen (2004). Concerning the learning-by-exporting and knowledge spillover explanations, see Herzer and D. (2006), Hausmann et al. (2007), and Agosin (2009). For general papers about the relationship between export diversification and growth, see Al-Marhubi (2000), Lederman and Maloney (2003), Newfarmer et al. (2009), or Cadot et al. (2012). For country- and region-specific analyses one might consider Alwang and Siegel (1994), Amin Gutiérrez de Piñeres and Ferrantino (1997), Taylor (2003), Sanguinetti et al. (2002), Petersson (2005), Beine and Coulombe (2007), Francis et al. (2007), Volpe Martincus and Carballo (2008), Naudé and Rossouw (2011), or El Hag and El Shazly (2012). For a general summary of the literature on export diversification with a focus on developing countries, one might look at Bonaglia and Fukasaku (2003) or Mejía (2011).

tential reverse causality between independent and dependent variables suggests using variables, which are averaged over longer time frames and lagged. But once one starts to average and lag, the number of observations quickly decreases, thereby reducing the interpretational power of results from conventional estimation techniques. The strength of BMA lies precisely in finding the true model, when faced with a smaller sample size and numerous potential determinants.

The literature on export diversification suggests a variety of determinants, yet the results in terms of significance vary widely. In addition, one finds substantial differences in the usage of control variables. For instance, consider a basic comparison of recent panel analyses, displayed in table 1. Notice the two final columns: the number of regressors varies between 7 – 23 (not counting the nonparametric analysis) and only two variables appear twice as significant factors. Thus, we conclude that model uncertainty presents a severe problem in assessing the determinants of export diversification.

To address this problem, the Bayesian Model Averaging (BMA) method presents itself as an ideal tool. We gather 43 potential long-run determinants of export diversification and analyze them in one cross-section sample of 89 countries. The purpose of the BMA technique is to consider all possible variable combinations (2^{43}) and reveal the true model. In order to best control for potential endogeneity, business cycles, external shocks, and measurement error, we average each independent variable from 1960 – 2000 in our main specification. Our measurement for export diversification as the dependent variable is the Herfindahl-Hirschman index (HHI) averaged from 2000 – 2010 for each country.

The results suggest that the most important variables in predicting long-term export diversification are the fraction of natural resource rents in GDP and the net primary enrollment rate. Further, population size and foreign direct investment play a secondary role. A major part of this result is the non-significance of many other potential determinants. For instance, neither trade-related aspects (e.g. regulatory trade freedom, trade openness) nor geographical remoteness (being an island or landlocked) nor a variety of macroeconomic factors seem to matter in predicting the long-run diversification level of a country's export basket.

The following section presents our methodology, followed by a description of our data in section 3. Section 4 discusses the Herfindahl-Hirschman index and potential determinants of export diversification, putting them into the context of previous works. Section 5 presents our

main results, followed by robustness checks in section 6. Finally, section 7 concludes and section 8 provides avenues for further research.

2 Methodology

A common problem of trying to assess the true long-run determinants of a macroeconomic variable is the combination of potential endogeneity, business cycles, and model uncertainty. For instance, whether richer countries diversify or diversification increases GDP is difficult to disentangle as both directions can plausibly be justified. Ideally, the researcher would like to have instruments that are strongly correlated with the independent variable, yet have no direct correlation with the dependent variable. Although imperfect, lagged values of the independent variable form the most suitable alternative in many cases. Adding in the task of smoothing out business cycles and containing the impact of measurement error (which can be of particular importance in some developing regions for instance), one might choose to average variables over 5 or sometimes 10 years. Consequently, if one uses both lagged and averaged values, the number of observations dwindles quickly in a panel data setting. Finally, the choice of potential explanatory variables is oftentimes open-ended, implying model uncertainty. In terms of export diversification, all of the above points become prevalent issues.

2.1 Endogeneity and Business Cycles

In order to cope with problems of endogeneity and business cycles, we choose to lag and average the explanatory variables over decades. In our main specification, we regress a country's average Herfindahl-Hirschman index (HHI) over the years 2000 – 2010 on the averaged values of each independent variable over the time period 1960 – 2000.[2] By using the HHI as a measurement for export diversification, we follow the vast majority of previous analyses (e.g. Imbs and Wacziarg, 2003, Parteka, 2010, Wiig and Kolstad, 2012, Agosin et al., 2012, or Cadot et al., 2011). Please also see section 4.1 for more information on why we choose the HHI.

There are several reasons for using this methodology. First, although the long-term trend of

[2]The volatility measurements of the interest rate and the inflation rate are taken as the variance over annual observations spanning the entire time period. Please also see sections 4.2 and 4.3 for details. We choose to start with the HHI from 2000 because earlier data covers significantly less countries.

a country's degree of export diversification is relatively stable, the data exhibits substantial year-to-year variation for some countries. This leads us to believe that either short-term fluctuations in export diversification levels or in macroeconomic variables severely affect the short-run value of diversification, or that the data exhibits measurement error. Since this paper focuses on the long-term aspects, averaging variables over decades alleviates both problems. Second, by using values from 1960 – 2000 to predict an outcome in 2000 – 2010, we firmly address the problem of reverse causality. Thus, our regressions take the general form of

$$\overline{HHI}_i^{2000-2010} = \alpha + \vec{\beta}\vec{X}_i^{1960-2000} + \epsilon_i, \tag{1}$$

where $\vec{X}_i^{1960-2000}$ contains all 43 potential determinants, discussed in detail below. Finally, after collecting data for the HHI and all independent variables, our main sample consists of 89 countries.

2.2 Bayesian Model Averaging

The Bayesian Model Averaging approach (BMA) is designed to specifically address model uncertainty in linear regression estimations. This becomes particularly important if the researcher is faced with a large set of potential explanatory variables, but only with a relatively limited number of observations. In this context, Raftery (1995) has shown that standard variable selection procedures can give very misleading results.

2.2.1 General BMA Framework

A Bayesian solution to model uncertainty involves averaging over all possible combinations of predictors (Leamer, 1978). Raftery et al. (1997) show that in the presence of predictors' uncertainty, BMA procedures provide better predictive performance than 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 various algorithms based on the Markov Chain Monte Carlo strategy.

With $\mathcal{M} = \{M_1, M_2, \ldots, M_{2^K}\}$ 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 $\pi(\theta_r|M_r)$, a likelihood $\pi(y|\theta_r, M_r)$, and a posterior $\pi(\theta_r|y, M_r)$. Using standard probability arguments, the posterior $\pi(\theta_r|y)$ is then given by

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

The BMA logic establishes to obtain results for every model under consideration and to average them. Further, the posterior model probability is given by

$$\pi(M_r|y) = \frac{\pi(y|M_r)\pi(M_r)}{\pi(y)} \tag{3}$$

with a marginal likelihood of

$$\pi(y|M_r) = \int \pi(y|\theta_r, M_r)\pi(\theta_r|M_r)d\theta_r \tag{4}$$

and $\pi(M_r)$ being the prior model probability. The posterior predictive density can then be obtained using a few basic rules of probability:

$$\pi(y^f|y) = \sum_{r=1}^{2^K} \left\{ \int \pi(y^f|y, \theta_r, M_r)\pi(\theta|y, M_r)d\theta_r \right\} \pi(M_r|y). \tag{5}$$

2.2.2 Practical BMA Issues

Two main issues remain with respect to the implementation of BMA: the integral in equation (4) is difficult to implement and the number of variable combinations can be enormous. These issues can be handled using the Markov Chain Monte Carlo Model Composition (MC^3), an algorithm adopted from the original mechanism developed by Madigan et al. (1995).

The MC^3 procedure is a mechanism for sampling over a model space $\mathcal{M}$, based on a Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). 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, 2003). A candidate model M^c is drawn randomly from the set of models, including (i) the current model

$M^{(s-1)}$, (ii) all models which delete one predictor from $M^{(s-1)}$, and (iii) all models which add one predictor to $M^{(s-1)}$. The acceptance probability has the following form

$$\alpha(M^{(s-1)}, M^c) = Min\left\{\frac{\pi(y|M^c)\pi(M^c)}{\pi(y|M^{(s-1)})\pi(M^{(s-1)})}, 1\right\}. \tag{6}$$

In general, BMA has a huge computational burden. As a consequence, it is a good idea to use the Normal linear model. That is, given $y = \beta_0 i_N + X_r\beta_r + \mu$, where y is an N-dimensional vector, i_N is an $N \times 1$ vector of ones, X_r is an $N \times K_r$ matrix of predictors, and $\mu \sim \mathcal{N}_N(0, \tau^{-1}I_N)$, then $\pi(\beta_r|\tau) \sim \mathcal{N}_{K_r}(0_{K_r}, \tau^{-1}(g_r(X_r'X_r))^{-1})$.

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, $\pi(\tau) \propto \frac{1}{\tau}$ and $\pi(\beta_0) \propto 1$. Fernandez et al. (2001) 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. Finally, Ley and Steel (2009) propose to use 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. This prior only requires 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.

In summary, we estimate multiple Bayesian normal linear models whose differences are given by the combination of predictors, where we choose the best models using a MC^3 procedure based on a Metropolis-Hastings algorithm. Finally, we average these models, obtaining posterior parameters and predictive distributions.

3 Descriptive Statistics

Tables 2 and 3 show summary statistics, data sources, and (if applicable) the specific calculation of each variable in our sample. We choose the Herfindahl-Hirschman index of export diversification (HHI) as a measurement for export diversification. We calculate the average of a country's annual HHI values between the years 2000 – 2010, in order to control for busi-

ness cycles, any short-term fluctuations, and potential measurement errors. Similarly, for every country we average each explanatory variable over the time frame 1960 – 2000.

Our sample covers a wide range of countries. For instance, consider the first political variable, the polityIV index. Although the mean score is about 2, our sample economies range from -9.5 (almost entirely autocratic) to +10 (totally democratic). Table 4 shows the HHI for all our 89 countries out of which we count 20 African, 18 Asian, 28 European, 7 North American, 3 Oceanian, and 13 South American or Caribbean nations. Countries are sorted by HHI, starting with the least diversified nations. We notice that African countries (such as Gabon, Mali, or Burkina Faso) are among the most concentrated, in addition to the Republic of Yemen, Bahrain, and Venezuela. European (with the exception of Norway) and North American economies on the other hand have a more diversified composition of exports. Table 4 also confirms the broad variety of available countries, covering nations at any stage of development.

Altogether, we are testing for 43 explanatory variables: 15 political, 11 macroeconomic, 8 cultural, and 9 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, meaning that only countries for which we have at least one observation for each of the 43 variables in the time frame 1960 – 2000 enter the sample. However, we also control for different time frames, as discussed in our robustness section below.

4 Variables

This section discusses our measurement for export diversification and its potential long-term determinants. Given the large number of potential factors (43), we only provide a brief explanation for the inclusion of a variable and refer to the respective previous papers for further details. 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 point out 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 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.

4.1 The HHI Index

We choose a classic representation of export diversification as our dependent variable: the Herfindahl-Hirschman index (HHI).[3] This index allows us to capture both the intensive and the extensive margin of diversification.[4]

The raw data for the HHI comes 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 2000 – 2010.[5] Assuming country i's exports in year t total US\$$X_{it}$ (with $1 \leq i \leq n$ and $1 \leq t \leq T$) and x_{ij} displays its exports of good j in US\$ (with $1 \leq j \leq m$), we use the common formula

$$HHI_{it} = \sum_{j=1}^{m} \left(\frac{x_{ijt}}{X_{it}} \right)^2 \tag{7}$$

to calculate country i's level of export diversification in year t. This procedure gives us a yearly index for the export diversification levels of 89 countries over the time frame 2000 – 2010. Averaging for every country then gives us the dependent variable:

$$\overline{HHI}_i^{2000-2010} = \sum_{t=2000}^{2010} \left(\frac{HHI_{it}}{11} \right). \tag{8}$$

Notice that higher values of the HHI signal a higher level of concentration. Thus, the lower the HHI score, the more diversified a country's export basket.

[3] Although there are other ways to measure export diversification – such as the Ogive, the Gini, or the Theil index – the HHI is used in a majority of analyses. Further, Ali et al. (1991) conclude that scores of the indices are generally very similar.

[4] 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 et al. (2011) for empirical analyses. Chaney (2008) provides a deeper theoretical foundation for the intensive and the extensive margin of international trade.

[5] Agosin et al. (2012) use three-digit level data, whereas Dennis and Shepherd (2011) use the number of product lines, thus focusing on the extensive margin of diversification.

4.2 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 $polityIV$ variable measures the level of democracy, ranging from -10 (strongly autocratic) to $+10$ (strongly democratic).[6] In addition, government size (gov) is measured as government share of real GDP per capita.[7] 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 various measurements: (i) $tradefreedom$, ranging from $0 - 100$, (ii) three tariff rates (weighted mean of all products, manufactured products, and primary products), (iii) the logarithm of the inflation rate ($loginfl$), and (iv) the volatility of the inflation rate ($volinfl$).[8] 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. Tariffs and inflation capture the business environment in an international context. Although tariffs do not directly affect exports, they do reflect the general regulatory attitude towards international trade. In addition, countries with high tariffs generally face higher tariffs themselves on the international market. Further, we include the volatility of the inflation rate as a proxy of the general stability of monetary policy. The intuition here is that fluctuations in surrounding market conditions may well discourage potential trade partners.

Another branch of variables considers a country's education system. Various studies hint that schooling plays a substantial role in export diversification, e.g. Mengistu (2009), Parteka and Tamberi (2011), or Agosin et al. (2012). To this end, we include five education variables: the duration of primary and secondary education, secondary and tertiary enrollment rates (gross %), and total primary enrollment (net %).[9] Including measurements for both the time of education

[6] We use variable polity2 from the Polity IV data set, designed to facilitate the use of the polity regime measure in time-series analyses.

[7] 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. (2012), although in the context of economic growth.

[8] Volatility is calculated as the variance of the inflation rate over the years 1960 – 2000.

[9] We also considered the education variables from Cohen and Soto (2007), but their inclusion would result in

and the 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 et al. (2012) employ average years of schooling from Barro and Lee (2000), whereas we distinguish between the duration of primary and secondary education, plus enrollment rates at different stages.

Finally, given the extraordinary role of petroleum in the world economy and the predominant role of OPEC countries (Organization of the Petroleum Exporting Countries) in this market, especially towards the end of the 20th century, we add a dummy variable for OPEC member countries.

4.3 Macroeconomic Factors

In terms of general macroeconomic conditions, we include openness to trade (*trade*), *exports* and *imports* separately, total natural resource rents as a percentage of GDP (*natres*), gross capital formation (*capital*), and foreign direct investment (fdi), all measured as a percentage of GDP. We include *natres* as a proxy for the importance of primary resources in the domestic economy.[10] This variable is particularly important given the discussion about the "resource-curse" (Sachs and Warner, 2001) in the context of export diversification and economic growth.

Further, we add the logarithm of GDP per capita, a net barter terms of trade index (year 2000 = 100), fuel exports (as % of merchandise exports), and the real interest rate to the list.[11] To proxy for general stability of the financial sector in an economy, we also include the variance of the interest rate over the respective time frame. Similar to the argument above regarding the volatility of inflation, large fluctuations in financial conditions may deter international trade relations and therefore affect export diversification. Although Agosin et al. (2012) include the variance of the exchange rate in their analysis, we prefer the volatility of interest rates for the following reason: a substantial number of countries switched currencies at one point or another

losing 18 sample countries. However, pure correlation of their education variables and the respective ones used here is at least 80 percent.

[10]The World Bank defines this variable as the sum of oil rents, natural gas rents, coal rents (hard and soft), mineral rents, and forest rents.

[11]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 et al. (2012), and Tadesse and Shukralla (2013).

during our main 40-year time frame from 1960 – 2000, which might introduce a bias in measuring the variance. For example, if Spain switches from the Peseta to the Euro, then we would incur a structural break in exchange rates (and in the variance), although not necessarily due to any kind of market fluctuation. Using the volatility of interest rates and inflation rates frees the analysis from this problem, yet captures the general stability of financial markets.

Finally, we include the investment share of real GDP per capita in our analysis, following Ben Hammouda et al. (2006) and Dennis and Shepherd (2011).

4.4 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.[12] We readily 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 whether a country applies the common law system. Although one might not immediately relate both of these variables to export diversification, their impact on institutions and growth have been shown, for instance in Acemoglu et al. (2001). Thus, if these variables affect other major macroeconomic variables, we also test for an effect on export diversification.

4.5 Geographical Factors

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

5 Results

All our estimations are performed with the R package (Team, 2011), using the BAS library (Feldkircher and Zeugner, 2012). We apply Bayesian Model Averaging by using the birth/death algorithm as the MC^3 sampler. The number of iteration draws to be sampled (ex burn-ins) is

[12] See Dennis and Shepherd (2011) and Parteka and Tamberi (2011) for instance.

1,200,000 and the number of burn-in draws for the MC^3 sampler is 200,0000. With the algorithm visiting 28,068 of 2^{43} potential models, the correlation between the analytical and MC^3 posterior model probabilities turn out to be over 0.99. This indicates a good performance of the algorithm (Fernandez et al., 2001).

5.1 Main Results

Table 5 shows our main BMA results. The two most important predictors of export diversification are total natural resources rents as a percentage of GDP (*natres*) and total net enrollment in primary education (*edu5*). The posterior inclusion probability (PIP) of both variables is remarkable with 1 and 0.96, meaning that they are included in 100 % and 96 % of the best models. Further, population size (*logpop*, PIP of 25 %) and foreign direct investment (*fdi*, PIP of 17 %) play a significant role, although of minor importance. After that, the PIPs drop substantially to under 2.5 %. This suggests that there are only up to four meaningful long-run predictors of export diversification. In fact, figure 1 shows a mean of the posterior model size distribution of only 2.6.

The most notable aspect of this result is not only the importance of natural resources and education, but even more so the insignificance of a variety of other variables. Prominent potential determinants of diversification, such as trade aspects (general trade freedom, tariffs, trade openness) or geographical variables (remoteness captured by island and landlocked) are assigned a PIP of under 0.01, meaning that they have virtually no effect on export diversification in the long-run.

Column 5 (Cond. Pos. Sign) allows us to conclude the sign of coefficients. For instance, a value of 1 means that a variable has a positive effect on a country's HHI in all models and thus decreases export diversification. Looking at *natres*, we find strong evidence that countries in which natural resources play a strong role in the domestic economy find it harder to diversify their exports.

Primary education levels, population size, and foreign direct investment on the other hand increase diversification levels in the export basket. In terms of education levels, it is interesting to see the enrollment percentage of primary education – not secondary or tertiary – to be highly important. Thus, a broad base of basic education seems to be important for export

diversification. This result confirms the general importance of schooling, as found in Agosin et al. (2012), who control for average years of schooling only as a measurement for education. As we are including five different aspects of education, we are able to conclude that it is especially the primary enrollment rate, which leads to export diversification over the long run. The argument for population size confirms findings from Parteka and Tamberi (2011) and seems intuitive: with more people comes the capacity to produce a bigger variety of products in the first place, which may then translate into more variety in exports. A small country will find it hard to diversify, simply because of available labor resources. Finally, we confirm Tadesse and Shukralla (2013) in their finding of the positive effect from foreign direct investment on export diversification. Since high levels of foreign capital could capture a relative openness to international financial markets, one may speculate whether financial openness also encourages export diversification.

Table 6 provides a deeper look at the main results by displaying the 10 top models from our BMA procedure. The best model carries a strong posterior model probability of 0.60, including natural resource percentage in GDP and primary enrollment only. This suggests that the long-term level of export diversification across our sample of 89 countries is determined mostly by these two variables. Altogether, the top three models account for 80% posterior probability, meaning that there is strong evidence of the true model consisting only of up to four factors, as these are the only variables included in the top three models. This conclusion is further strengthened by the results in figure 1, suggesting a posterior model size distribution of 2.6. Further, the probability of models with 2 to 4 variables is approximately 0.95.

Finally, figure 2 depicts the posterior probability density functions of the coefficients associated with the most important variables. For instance, we notice that the density function of the coefficient associated with natural resource abundance in GDP is always positive. On the other hand, the density functions of primary enrollment, population size, and foreign investment are negative.

5.2 Notable Country Examples

We now turn to a brief overview of some countries with noteworthy levels of export diversification and their values of the most important variables. Specifically, we are looking at the ten most concentrated and the ten most diversified export baskets in table 7. Among the least diversified

countries, we find the Republic of Yemen, Gabon, Venezuela, and Bahrain – all with natural resource rents amounting to over 25 % of GDP. This stands in stark contrast to the countries with the highest diversification rates, where natural resources play minor roles. In general, the diversified economies are mostly marked by low dependence on natural resources, sizeable populations, and high primary enrollment rates. For the countries with concentrated export baskets, at least one of these aspects does not hold.

For instance, Mali and Burkina Faso share low dependence on natural resources and have modest population sizes, but also abysmal primary enrollment rates of under 25 %. China on the other hand relies on natural resources to a more substantial degree – although the country is far from the likes of Gabon and Bahrain – but seems to be able to compensate by an enormous population size and primary enrollment rates of over 96 % on average. Other countries such as Italy, Poland, the United States, France, or Spain combine the entire recipe for export diversification: low natural resource dependence, sizeable population, and primary enrollment rates well over 90 %. However, we can see that also relatively smaller countries like Austria or Denmark find it possible to diversify their exports exceptionally well by standing out in the other categories. Finally, we can see no notable pattern in comparing FDI levels among the least and the most diversified economies.

6 Robustness Checks

This section presents alternative specifications, addressing potential weaknesses of our main version. Specifically, we control for a potential endogeneity bias and our averaging period.

6.1 Reverse Causality

Our first robustness check considers a potential reverse causality between any independent variable and our measurement for export diversification, the HHI. Time frames briefly overlap in our main results with the HHI averaged from 2000 – 2010 regressed on averages from 1960 – 2000 of each independent variable. To create a gap between both types of variables, we re-run our BMA model using the average HHI from 2005 – 2010. Using this 6-year window still gives us enough time to reasonably smooth over short-run fluctuations, which seems important espe-

cially around the financial crisis in 2007. This specification strengthens our argument against reverse causality as the outcome now occurs between 5 and 50 years after the observation of the independent variables.

With a sample of 88 countries we only lose Papua New Guinea due to missing observations of the HHI and table 8 displays our results from using the 2005 – 2010 window for the HHI. A quick comparison to our main results shows a strong resemblance to our main results with the same four variables at the top.

6.2 Averaging Method

With our explanatory variables being averaged over 40 years in our main results, one may suspect a potential selection bias. For instance, reliability and availability of many variables could be different in developing versus developed nations. This means that a variable might be averaged over very few observations in one country, but over the entire time frame in another country, thereby affecting the comparability of observations. To address this problem, we re-estimate our BMA model by averaging all explanatory variables only from 1990 – 2000. Averaging over 10 years strengthens the comparability of observations, yet still allows us to reduce potential problems from short-run fluctuations, measurement errors, or exogenous shocks.

Table 9 shows the respective results and we notice a barely reduced sample size of 83 countries. This reduction stems from the loss of 6 countries, which do not have observations for the time period 1990 – 2000 for at least one of the 43 explanatory variables.[13] As before, the fraction of natural resources in GDP and primary education are the dominating factors with a dummy for the African continent coming in third. Notice that the posterior inclusion probability of primary enrollment rates is now only 0.54 as opposed to values over 0.96 before. This suggests either that averaging primary enrollment over 40 years creates a bias or that primary education affects export diversification in the *very* long run. Additionally, the dummy for Africa is gaining importance over time with a PIP of 0.22. Compared with our main results, where being an African country is a non-factor with a PIP of 0.02, this difference is remarkable. Also, population size and foreign direct investment lose their impact in this robustness check. Natural resources and primary education remain the most important determinants, underpinning their

[13] We are losing Ghana, Sri Lanka, Thailand, Tunisia, Uganda, and Zimbabwe.

significance in predicting export diversification levels in the long run.

7 Conclusions

This paper seeks to provide a deeper understanding of the long-term determinants of export diversification. With the lack of a comprehensive theoretical foundation and a variety of potential empirical determinants previously suggested, it is difficult to find the true model explaining a country's level of export diversification. However, using the Bayesian Model Averaging (BMA) method allows us to get to the grain of model uncertainty.

Our main specification comprises 89 countries and considers 43 potential determinants of export diversification. This means that there are 8,796,093,022,208 (2^{43}) variable combinations under which the true model is hidden. Our results suggest that only four variables play a decisive role in predicting export diversification levels: (*i*) the importance of natural resource rents in an economy (100 % posterior inclusion probability in the true model), (*ii*) the total net primary enrollment rate (96 %), (*iii*) population size (25 %), and (*iv*) foreign direct investment levels (17 %). Especially primary education rates and foreign direct investment levels offer itself for potential policy recommendations. If, as previous works suggest, a diversified export basket may lower income volatility, generate knowledge spillovers, and open doors to a smoother development path, then our paper provides additional support for the importance of basic education levels and providing conditions for international capital to enter.

Beyond the significant factors, our results offer equally important conclusions regarding the non-significance of many variables. Neither trade-related (e.g. trade openness, trade freedom, or terms of trade) nor political factors (e.g. degree of democracy, size and effectiveness of government) are playing a role in the long-term determination of export diversification. Similarly, macroeconomic aspects (such as interest rates, inflation rates, or their volatilities) are also insignificant. The same holds true for the general remoteness of a country, measured with continental dummies and binary variables for islands and landlocked economies. We believe that these results are an important step towards understanding the true roots of export diversification.

8 Further Research

Regarding the limitations of this paper, we believe that there are two main technical weaknesses, owed to our econometric methodology. The Bayesian Model Averaging methodology neither allows us to use panel data nor to restrict the dependent variable (the Herfindahl-Hirschman index) between zero and one. The first point implies that we cannot control for any unobserved country-specific aspects. Although the inclusion of 43 explanatory variables substantially alleviates this problem, it cannot eliminate it. The second point represents a technical problem owed to the current econometric state of the BMA technique. In practice, we may only incur minor estimation glitches, but the pure econometrician surely notices this shortcoming. Both challenges are left for future research.

Finally, our findings of the mentioned four important long-run determinants of export diversification leave room for more detailed analyses. Specifically, exploring the exact channels and assessing the quantitative implications on diversification levels may provide fruitful avenues for future research.

References

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5):1369–1401.

Agosin, M. (2006). Trade and growth: Why asia grows faster than latin america. *IDB Publications*.

Agosin, M. (2009). Export diversification and growth in emerging economies. *CEPAL Review*, 97:115–131.

Agosin, M., Alvarez, R., and Bravo-Ortega, C. (2012). Determinants of export diversification around the world: 1962–2000. *The World Economy*.

Al-Marhubi, F. (2000). Export diversification and growth: an empirical investigation. *Applied Economics Letters*, 7(9):559–562.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2):155–194.

Ali, R., Alwang, J., and Siegel, P. (1991). *Is export diversification the best way to achieve export growth and stability?: a look at three African countries*, volume 729. World Bank Publications.

Alwang, J. and Siegel, P. (1994). Portfolio models and planning for export diversification: Malawi, tanzania and zimbabwe. *The Journal of Development Studies*, 30(2):405–422.

Amin Gutiérrez de Piñeres, S. and Ferrantino, M. (1997). Export diversification and structural dynamics in the growth process: The case of chile. *Journal of Development Economics*, 52(2):375–391.

Amurgo-Pacheco, A. (2008). *Patterns of export diversification in developing countries: intensive and extensive margins*, volume 4473. World Bank Publications.

Barro, R. and Lee, J. (2000). International data on educational attainment updates and implications. *NBER Working Paper Series*, 7911.

Bebczuk, R. and Berrettoni, N. (2006). Explaining export diversification: an empirical analysis. *Documento de Trabajo*, (65).

Beine, M. and Coulombe, S. (2007). Economic integration and the diversification of regional exports: evidence from the canadian–us free trade agreement. *Journal of Economic Geography*, 7(1):93–111.

Ben Hammouda, H., Karingi, S., Njuguna, A., and Sadni Jallab, M. (2006). Diversification: towards a new paradigm for africa's development. *ATPC work in progress*, (35).

Bonaglia, F. and Fukasaku, K. (2003). Export diversification in low-income countries: An international challenge after doha. *OECD Development Centre Working Papers*.

Breton, T. R. (2012). Penn world table 7.0: Are the data flawed? *Economics Letters*, 117(1):208 – 210.

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

Cadot, O., Carrère, C., and Strauss-Kahn, V. (2012). Trade diversification, income, and growth: What do we know? *Journal of Economic Surveys*.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, 98(4):1707–1721.

Cohen, D. and Soto, M. (2007). Growth and human capital: good data, good results. *Journal of Economic Growth*, 12(1):51–76.

Dennis, A. and Shepherd, B. (2011). Trade facilitation and export diversification. *The World Economy*, 34(1):101–122.

El Hag, S. and El Shazly, M. (2012). Oil dependency, export diversification and economic growth in the arab gulf states. *European Journal of Social Sciences*, 29(3):397–404.

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

Fernandez, C., Ley, E., and Steel, M. (2001). Benchmark priors for bayesian model averaging. *Journal of Econometrics*, 100:381–427.

Francis, B., Iyare, S., and Lorde, T. (2007). Agricultural export-diversification and economic growth in caribbean countries: Cointegration and error-correction models. *The International Trade Journal*, 21(3):229–256.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their application. *Biometrika*, 57:97–109.

Hausmann, R., Hwang, J., and Rodrik, D. (2007). What you export matters. *Journal of Economic Growth*, 12(1):1–25.

Herzer, D. and D., F. N.-L. (2006). What does export diversification do for growth? an econometric analysis. *Applied Economics*, 38(15):1825–1838.

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

Jansen, M. (2004). Income volatility in small and developing economies: export concentration matters. *Discussion Paper, World Trade Organization*, (3).

Johnson, S., Larson, W., Papageorgiou, C., and Subramanian, A. (2012). Is newer better? penn world table revisions and their impact on growth estimates. *Journal of Monetary Economics*, (0):–.

Koop, G. (2003). *Bayesian Econometrics*. Jhon Wiley & Sons Ltd.

Leamer, E. (1978). *Specification Searches*. New York: Weley, New York.

Lederman, D. and Maloney, W. (2003). Trade structure and growth. *World Bank Policy Research Working Paper*, (3025).

Ley, E. and Steel, M. (2009). On the effect of prior assumptions in bayesian model averaging with applications to growth regressions. *Journal of Applied Econometrics*, 24(4):651–674.

Madigan, D., j. York, and Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63(2):215–232.

Mejía, J. (2011). *Export Diversification and Economic Growth: An Analysis of Colombia's Export Competitiveness in the European Union's Market*. Physica-Verlag HD.

Mengistu, A. (2009). 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):1–2.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys*, 21:1087–1092.

Naudé, W. and Rossouw, R. (2011). Export diversification and economic performance: evidence from brazil, china, india and south africa. *Economic Change and Restructuring*, 44(1):99–134.

Newfarmer, R., Shaw, W., and Walkenhorst, P. (2009). *Breaking into new markets: Emerging lessons for export diversification*. World Bank Publications.

Parteka, A. (2010). Employment and export specialisation along the development path: some robust evidence. *Review of World Economics*, 145(4):615–640.

Parteka, A. and Tamberi, M. (2008). Determinants of export diversification: an empirical investigation. *Dipartimento di Economia Quaderno di Ricerca*, (327).

Parteka, A. and Tamberi, M. (2011). Export diversification and development-empirical assessment. *Universita'Politecnica delle Marche Dipartimento di Scienze Economiche e Sociali Working Paper*, (359).

Petersson, L. (2005). Export diversification and intra-industry trade in south africa. *South African Journal of Economics*, 73(4):785–802.

Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25:111–163.

Raftery, A., Madigan, D., and Hoeting, J. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191.

Sachs, J. and Warner, A. (2001). The curse of natural resources. *European Economic Review*, 45(4):827–838.

Sanguinetti, P., Pantano, J., and Arim, R. (2002). Trade liberalization and exports/imports diversification in argentina: the role of tariff preferences and economies of scale. Department of Economics Working Papers 015, Universidad Torcuato Di Tella.

Tadesse, B. and Shukralla, E. (2013). The impact of foreign direct investment on horizontal export diversification: empirical evidence. *Applied Economics*, 45(2):141–159.

Taylor, T. (2003). Export diversification in latin america and the caribbean. *The International Trade Journal*, 17(2):101–128.

Team, R. D. C. (2011). *R: A Language and Environment for Statistical Computing. R. Version 2.13.0.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Volpe Martincus, C. and Carballo, J. (2008). 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):89–106.

Wiig, A. and Kolstad, I. (2012). If diversification is good, why don't countries diversify more? the political economy of diversification in resource-rich countries. *Energy Policy*, 40(0):196 – 203. Strategic Choices for Renewable Energy Investment.

Figure 1: Bayesian Model Averaging: Prior and Posterior Model Size

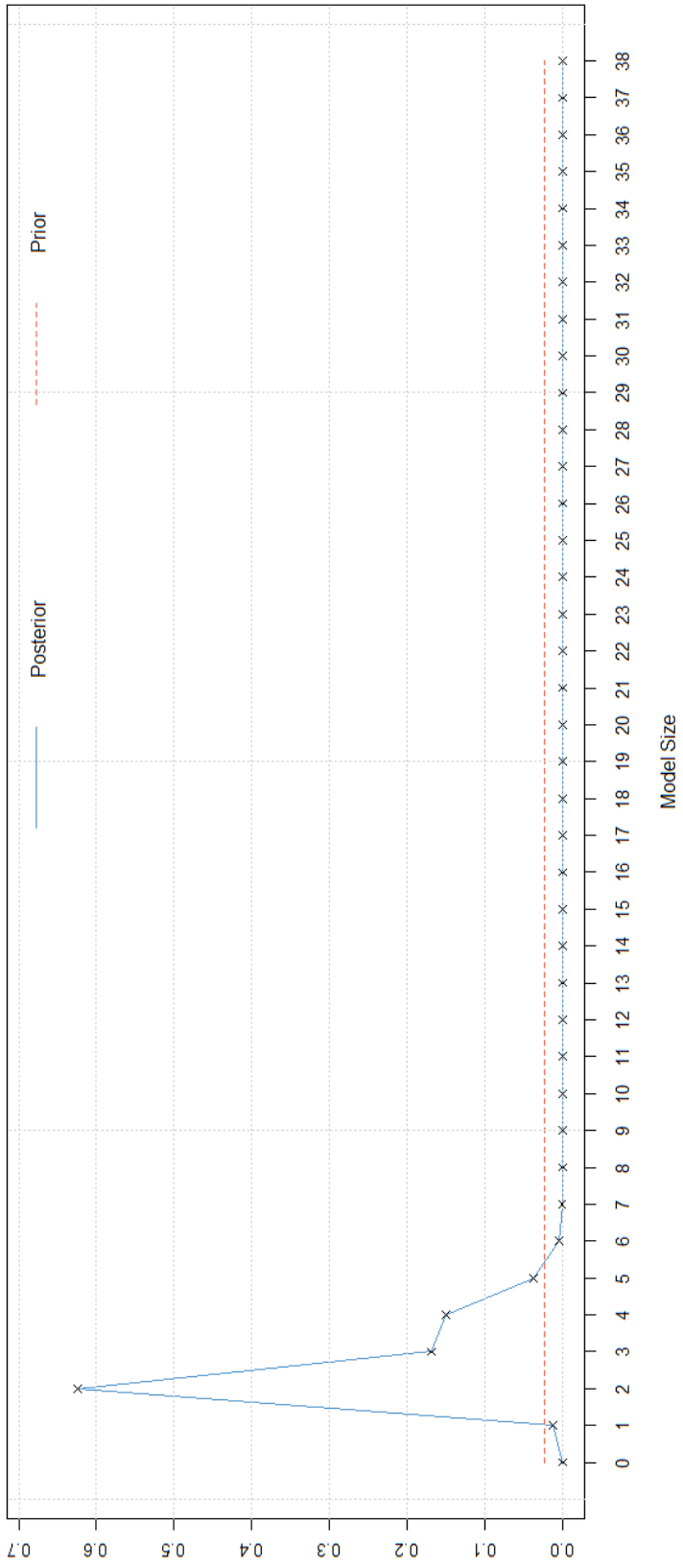


Figure 2: Bayesian Model Averaging: Posterior Probability Density of the most important coefficients

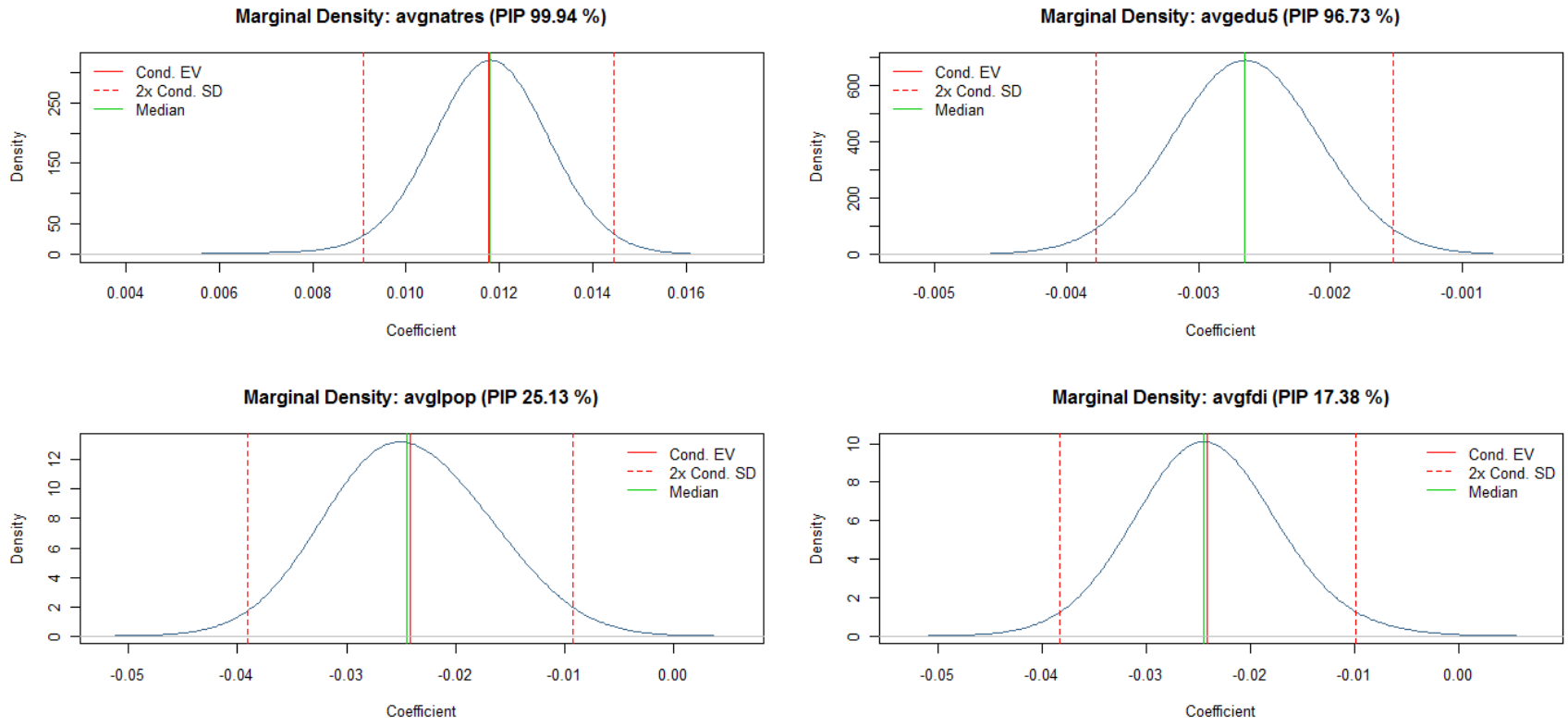


Table 1: Recent Panel Studies on the Determinants of Export Diversification.

Paper	Authors	Dependent Variable	# of Regressors	Empirical Method	Significant Determinants
Determinants of Export Diversification Around the World: 1962 – 2000	Agosin et al. (2012)	HHI, Gini, Theil	7	GMM system estimator	Remoteness (-), trade openness (-), schooling (+), human capital interacted with terms of trade
Trade Facilitation and ED	Dennis and Shepherd (2011)	# of exports	23	Poisson Estimation, IV Poisson, 2 SLS	Costs of export, market entry, and international transport (all -)
Employment and export specialisation along the development path: some robust evidence	Parteka (2010)	HHI, Gini, Theil	2	semiparametric 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: towards 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 (-)

Table 2: Summary statistics. All values represent country averages from 1960 – 2000. The HHI is averaged from 2000 – 2010. volinfl and volint represent the variance of inflation and interest rates over the time frame 1960 – 2000.

Variable	Variable Name	Mean	(Std. Dev.)	Min.	Max.	N
Herfindahl-Hirschman Index	HHI	0.101	(0.14)	0.004	0.648	89
Political Factors						
Polity IV index	polityIV	1.994	(6.324)	-9.533	10	89
Government size	gov	17.767	(8.4)	5.256	53.484	89
Government effectiveness	goveff	0.239	(0.918)	-1.07	2.053	89
Trade freedom	tradefreedom	62.213	(14.756)	14	81.567	89
Tariff, weighted mean all products	tariff1	11.337	(8.152)	0	50.428	89
Tariff, weighted mean manufactured products	tariff2	11.497	(9.284)	0	59.275	89
Tariff, weighted mean primary products	tariff3	11.485	(9.048)	0	49.169	89
Inflation rate	loginfl	2.679	(1.237)	0.813	6.391	89
Variance of inflation rate	volinfl	5.165	(3.116)	0.877	15.056	89
Duration secondary education	edu1	6.387	(0.896)	4	9	89
Duration primary education	edu2	5.567	(0.916)	3	7	89
Secondary school enrollment, gross %	edu3	62.164	(31.658)	3.844	140.155	89
Tertiary school enrollment, gross %	edu4	18.419	(15.462)	0.279	78.350	89
Primary school enrollment, net %	edu5	84.717	(17.519)	20.959	99.708	89
OPEC membership	opec	0.011	(0.106)	0	1	89
Macroeconomic Factors						
Trade openness	trade	63.261	(33.702)	13.846	181.011	89
Value of exports	exports	29.797	(17.229)	6.313	95.053	89
Value of imports	imports	33.464	(17.142)	7.23	85.958	89
Natural resource rents in % of GDP	natres	5.829	(7.796)	0.024	37.371	89
Gross capital formation as % of GDP	capital	22.914	(4.691)	10.244	35.178	89
Foreign direct investment	fdi	1.771	(1.444)	-0.067	6.055	89
GDP per capita	loggdppc	7.625	(1.502)	4.832	10.356	89
Terms of trade	tot	106.596	(19.378)	82.89	180.784	89
Fuel exports	fuel	11.722	(19.49)	0.001	91.551	89
Real interest rate	interest	6.964	(12.032)	-48.442	64.594	89
Variance of real interest rate	volint	3.915	(1.816)	-1.292	9.516	89
Investment share of GDP	invest	21.504	(9.938)	3.339	57.745	89
Cultural Factors						
Population	logpop	16.231	(1.496)	12.8	20.695	89
Former British colony	british	0.27	(0.446)	0	1	89
Former Dutch colony	dutch	0.067	(0.252)	0	1	89
Former French colony	french	0.191	(0.395)	0	1	89
Former Portuguese colony	portuguese	0.045	(0.208)	0	1	89
Former Spanish colony	spanish	0.191	(0.395)	0	1	89
Language fractionalization	language	0.353	(0.288)	0.002	0.923	89
English common law system	commonlaw	0.236	(0.427)	0	1	89
Geographical Factors						
Africa	africa	0.225	(0.42)	0	1	89
Asia	asia	0.202	(0.404)	0	1	89
Europe	europe	0.315	(0.467)	0	1	89
North America	namerica	0.079	(0.271)	0	1	89
Oceania	oceania	0.034	(0.181)	0	1	89
South America and the Caribbean	smamerica	0.146	(0.355)	0	1	89
Island	island	0.169	(0.376)	0	1	89
Landlocked	landlocked	0.191	(0.395)	0	1	89

Table 3: Data Sources and Calculations

Variable Name	**Source**	**Calculation**
HHI	United Nations	ComTrade dataset, 6-digit level, see paper for details
Political Factors		
polityIV	Polity IV	Variable polity2 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)
tariff1	World Bank	tariff rate applied, weighted mean, weighted mean, all products in %
tariff2	World Bank	tariff rate applied, weighted mean, manufactured products in %
tariff3	World Bank	tariff rate applied, weighted mean, primary products in %
loginfl	Global Development Network Growth Database	ln(average inflation rate)
volinfl	own calculation	ln(variance of inflation rate over respective time period)
edu1	World Bank	duration of secondary education in years
edu2	World Bank	duration of primary education in years
edu3	World Bank	school enrollment secondary, gross %
edu4	World Bank	school enrollment tertiary, gross %
edu5	World Bank	total enrollment primary, net %
opec		dummy for OPEC member in the major part between 1960-2000

Table 3 continued: Data Sources and Calculations

Variable	Source	Calculation
Macroeconomic Factors		
trade	World Bank	(exports + imports)/GDP
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
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)
interest	World Bank	real interest rate (%)
volint	own calculation	ln(variance of interest rate over respective time period)
invest	PWT 6.3	investment share of real gdp per capita
Cultural Factors		
logpop	World Bank	ln(total population)
british, dutch, french, portuguese, spanish		5 dummies for former colonies
language	Alesina et al. (2003)	language fractionalization
commonlaw		dummy for English common law system
Geographical Factors		
africa, asia, europe, namerica, oceania, smamerica		6 continental dummies for Africa, Asia, Europe, North America, Oceania, South America and Caribbean
island, landlocked		2 dummies for whether a country is an island/landlocked

Table 4: Countries by average HHI from 2000 – 2010.

Country	HHI	Country	HHI	Country	HHI
Africa		**Europe**		**Oceania**	
Gabon	0.628	Norway	0.211	Papua New Guinea	0.169
Mali	0.473	Belarus	0.081	Australia	0.047
Burkina Faso	0.424	Albania	0.050	New Zealand	0.016
Ghana	0.314	Lithuania	0.044		
Mozambique	0.302	Georgia	0.041	**South America and the Caribbean**	
Algeria	0.298	Cyprus	0.039		
Cameroon	0.231	Netherlands	0.032	Venezuela, RB	0.528
Zambia	0.224	Moldova	0.028	Jamaica	0.271
Malawi	0.211	Latvia	0.023	Ecuador	0.265
Ethiopia	0.163	Estonia	0.022	Paraguay	0.145
Egypt, Arab Rep.	0.100	Finland	0.020	Bolivia	0.142
Zimbabwe	0.088	Hungary	0.019	Guyana	0.122
Tanzania	0.078	Bulgaria	0.019	Colombia	0.080
Kenya	0.074	Switzerland	0.017	Peru	0.076
Mauritius	0.070	Greece	0.016	Costa Rica	0.058
Uganda	0.056	Portugal	0.015	Guatemala	0.036
Madagascar	0.044	United Kingdom	0.014	Uruguay	0.036
Tunisia	0.024	Sweden	0.013	Argentina	0.031
Morocco	0.022	Romania	0.011	Brazil	0.016
South Africa	0.020	Slovenia	0.010		
		Denmark	0.010		
Asia		Germany	0.009		
		Spain	0.009		
Yemen, Rep.	0.648	Austria	0.007		
Bahrain	0.460	Czech Republic	0.007		
Israel	0.179	Poland	0.007		
Kyrgyz Republic	0.162	France	0.007		
Russian Federation	0.141	Italy	0.004		
Philippines	0.077				
Vietnam	0.048	**North America**			
Nepal	0.045				
Bangladesh	0.041	Trinidad and Tobago	0.135		
India	0.031	Honduras	0.062		
Malaysia	0.026	Dominican Republic	0.038		
Indonesia	0.025	Panama	0.038		
Korea, Rep.	0.022	Mexico	0.025		
Sri Lanka	0.020	Canada	0.024		
Jordan	0.016	United States	0.007		
Japan	0.013				
Thailand	0.011				
China	0.007				

Table 5: Bayesian Model Averaging: Main results averaging predictors from 1960 to 2000 and HHI from 2000 to 2010 (89 countries).

Variable	PIP	Post Mean	Post SD	Cond.Pos.Sign
natres	0.9989	1.17E-02	1.41E-03	1.0000
edu5	0.9587	-2.54E-03	7.65E-04	0.0000
logpop	0.2671	-6.45E-03	1.14E-02	0.0000
fdi	0.1848	-4.47E-03	9.88E-03	0.0000
africa	0.0240	1.92E-03	1.33E-02	1.0000
fuel	0.0229	3.93E-05	2.92E-04	1.0000
gov	0.0174	4.23E-05	3.57E-04	1.0000
tradefreedom	0.0090	1.20E-05	1.45E-04	0.9958
namerica	0.0085	-4.68E-04	5.96E-03	0.0000
tariff2	0.0079	-1.46E-05	1.92E-04	0.0006
volint	0.0077	6.91E-05	9.32E-04	1.0000
landlocked	0.0074	3.05E-04	4.25E-03	1.0000
island	0.0055	-1.74E-04	2.99E-03	0.0000
dutch	0.0051	-2.73E-04	4.68E-03	0.0000
exports	0.0048	-5.16E-06	1.89E-04	0.3740
imports	0.0046	7.02E-06	1.90E-04	0.9113
edu3	0.0044	1.06E-06	4.44E-05	0.8562
trade	0.0042	7.09E-07	1.21E-04	0.6063
invest	0.0041	-6.06E-06	1.37E-04	0.0000
tariff1	0.0040	-6.29E-06	1.30E-04	0.0097
edu4	0.0040	2.25E-06	6.30E-05	0.9476
smamerica	0.0038	1.02E-04	2.36E-03	1.0000
british	0.0032	-2.41E-05	1.26E-03	0.0251
french	0.0031	6.63E-05	1.95E-03	1.0000
loggdppc	0.0029	-7.47E-07	6.75E-04	0.6275
spanish	0.0028	3.14E-05	1.54E-03	0.8377
edu2	0.0028	-5.84E-07	6.26E-04	0.6443
commonlaw	0.0028	-2.23E-05	1.26E-03	0.0242
language	0.0028	1.08E-04	3.27E-03	0.9952
goveff	0.0026	-2.70E-05	9.85E-04	0.0233
asia	0.0026	-1.25E-05	1.44E-03	0.3971
interest	0.0025	7.74E-07	4.20E-05	0.9974
capital	0.0025	3.65E-06	1.52E-04	0.9357
opec	0.0024	1.11E-04	5.17E-03	0.9637
volinfl	0.0023	4.20E-06	1.73E-04	0.9929
tariff3	0.0023	-6.40E-07	5.59E-05	0.2025
loginfl	0.0022	3.75E-06	3.92E-04	0.9252
polityIV	0.0018	8.97E-07	7.67E-05	0.9131
tot	0.0017	2.49E-07	2.29E-05	0.7441
infl	0.0017	-4.39E-08	4.10E-06	0.1717
europe	0.0017	5.44E-06	1.00E-03	0.8741
edu1	0.0017	1.99E-06	4.51E-04	0.8832
portuguese	0.0016	-1.87E-05	1.96E-03	0.0840

Table 6: Bayesian Model Averaging: Variables included in 10 top models averaging predictors from 1960 to 2000 and HHI from 2000 to 2010 (89 countries).

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
natres	1	1	1	1	1	1	1	1	1	1
edu5	1	1	1	0	0	1	1	1	1	1
logpop	0	1	1	0	0	0	1	0	1	0
fdi	0	1	0	0	0	1	1	0	1	0
africa	0	0	0	1	0	0	0	0	0	0
fuel	0	0	0	0	0	0	1	0	0	1
gov	0	0	0	0	0	0	0	0	1	0
tradefreedom	0	0	0	0	0	0	0	1	0	0
namerica	0	0	0	0	0	0	0	0	0	0
tariff2	0	0	0	0	0	0	0	0	0	0
volint	0	0	0	0	0	0	0	0	0	0
landlocked	0	0	0	0	0	0	0	0	0	0
island	0	0	0	0	0	0	0	0	0	0
dutch	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
edu3	0	0	0	0	0	0	0	0	0	0
trade	0	0	0	0	0	0	0	0	0	0
invest	0	0	0	0	0	0	0	0	0	0
tariff1	0	0	0	0	0	0	0	0	0	0
edu4	0	0	0	0	0	0	0	0	0	0
smamerica	0	0	0	0	0	0	0	0	0	0
british	0	0	0	0	0	0	0	0	0	0
french	0	0	0	0	0	0	0	0	0	0
loggdppc	0	0	0	0	0	0	0	0	0	0
spanish	0	0	0	0	0	0	0	0	0	0
edu2	0	0	0	0	0	0	0	0	0	0
commonlaw	0	0	0	0	0	0	0	0	0	0
language	0	0	0	0	0	0	0	0	0	0
goveff	0	0	0	0	0	0	0	0	0	0
asia	0	0	0	0	0	0	0	0	0	0
interest	0	0	0	0	0	0	0	0	0	0
capital	0	0	0	0	0	0	0	0	0	0
opec	0	0	0	0	0	0	0	0	0	0
volinfl	0	0	0	0	0	0	0	0	0	0
tariff3	0	0	0	0	0	0	0	0	0	0
loginfl	0	0	0	0	0	0	0	0	0	0
polityIV	0	0	0	0	0	0	0	0	0	0
tot	0	0	0	0	0	0	0	0	0	0
infl	0	0	0	0	0	0	0	0	0	0
europe	0	0	0	0	0	0	0	0	0	0
edu1	0	0	0	0	0	0	0	0	0	0
portuguese	0	0	0	0	0	0	0	0	0	0
PMP (Exact)	0.6024	0.1227	0.0740	0.0128	0.0118	0.0082	0.0063	0.0062	0.0060	0.0059
PMP (MCMC)	0.6016	0.1204	0.0741	0.0139	0.0135	0.0083	0.0057	0.0052	0.0053	0.0056

Table 7: The 10 Least and the 10 Most Diversified Export Baskets.

Country	HHI	Nat. Res. Rents (% of GDP)	Population in Mill.	Prim. Enrollment Rate	FDI (% of GDP)
10 Least Diversified Export Baskets					
Yemen, Rep.	0.648	27.3	9.4	56.7	1.4
Gabon	0.628	37.4	0.7	91.8	-0.1
Venezuela, RB	0.528	27.5	15.3	81.6	1.0
Mali	0.473	2.4	7.5	22.3	0.5
Bahrain	0.460	35.3	0.4	90.7	4.3
Burkina Faso	0.424	4.2	7.7	21.0	0.2
Ghana	0.314	3.8	11.9	63.2	1.0
Mozambique	0.302	5.6	12.1	43.0	1.4
Algeria	0.298	17.9	19.6	84.1	0.5
Jamaica	0.271	6.9	2.1	95.6	1.5
10 Most Diversified Export Baskets					
Italy	0.004	0.2	55.2	99.0	0.3
Poland	0.007	1.1	35.2	97.4	1.7
Czech Republic	0.007	0.4	10.1	91.5	4.6
Austria	0.007	0.7	7.6	88.6	0.7
China	0.007	7.4	972.3	96.1	2.5
United States	0.007	2.1	228.6	94.4	0.8
France	0.007	0.2	54.9	98.4	0.9
Germany	0.009	0.3	78.4	87.4	0.7
Spain	0.009	0.2	36.3	99.2	1.5
Denmark	0.010	0.6	5.0	97.9	1.8

Note: HHI values are averages from 2000 – 2010. All other values are averages over the time frame 1960 – 2000.

Table 8: Bayesian Model Averaging: Results from averaging predictors from 1960 to 2000 and HHI from 2005 to 2010 (88 countries).

Variable	PIP	Post Mean	Post SD	Cond. Pos. Sign.
natres	1.0000	1.34E-02	1.39E-03	1.0000
edu5	0.9690	-2.71E-03	7.57E-04	0.0000
logpop	0.1820	-4.12E-03	9.35E-03	0.0000
fdi	0.0716	-1.64E-03	6.28E-03	0.0000
gov	0.0119	2.75E-05	2.84E-04	1.0000
tradefreedom	0.0110	1.58E-05	1.71E-04	0.9972
tariff2	0.0109	-2.31E-05	2.54E-04	0.0000
africa	0.0094	6.26E-04	7.53E-03	0.9768
namerica	0.0074	-4.27E-04	5.86E-03	0.0000
fuel	0.0072	9.54E-06	1.33E-04	1.0000
landlocked	0.0071	3.07E-04	4.38E-03	1.0000
dutch	0.0061	-3.79E-04	5.72E-03	0.0000
volint	0.0057	4.93E-05	7.92E-04	1.0000
tariff1	0.0047	-7.62E-06	1.58E-04	0.0256
exports	0.0047	-7.03E-06	1.88E-04	0.3023
trade	0.0045	-5.51E-07	9.75E-05	0.5607
avgedu4	0.0044	3.70E-06	7.94E-05	0.9697
polityIV	0.0044	1.09E-05	2.10E-04	0.9981
imports	0.0042	4.22E-06	1.63E-04	0.8331
loggdppc	0.0039	2.25E-05	7.85E-04	0.8909
smamerica	0.0036	1.01E-04	2.39E-03	1.0000
asia	0.0036	-8.09E-05	2.16E-03	0.1104
island	0.0036	-8.67E-05	2.20E-03	0.0000
capital	0.0035	8.49E-06	2.15E-04	0.9755
edu3	0.0034	-5.21E-08	4.25E-05	0.7185
invest	0.0030	-3.01E-06	1.07E-04	0.0204
british	0.0030	-3.90E-05	1.47E-03	0.0062
language	0.0029	8.57E-05	3.14E-03	0.9936
opec	0.0027	1.51E-04	5.70E-03	0.9938
commonlaw	0.0025	-2.57E-05	1.28E-03	0.0218
interest	0.0024	1.08E-06	4.57E-05	1.0000
europe	0.0022	8.02E-06	1.23E-03	0.7231
edu1	0.0022	-2.53E-07	5.41E-04	0.1215
goveff	0.0021	-1.61E-05	8.64E-04	0.0212
portuguese	0.0021	-9.22E-06	2.26E-03	0.0579
edu2	0.0020	7.16E-06	5.38E-04	0.9296
tot	0.0020	4.48E-07	2.70E-05	1.0000
french	0.0020	3.90E-05	1.57E-03	1.0000
volinfl	0.0019	4.24E-06	1.94E-04	0.9899
loginfl	0.0017	2.80E-06	4.36E-04	0.8783
spanish	0.0017	1.56E-05	1.16E-03	0.8717
tariff3	0.0016	-1.33E-07	4.89E-05	0.3655

Table 9: Bayesian Model Averaging: results from averaging predictors from 1990 to 2000 and HHI from 2000 to 2010 (83 countries).

Variable	PIP	Post Mean	Post SD	Cond. Pos. Sign.
natres	0.9966	1.25E-02	1.73E-03	1.0000
edu5	0.5419	-1.41E-03	1.38E-03	0.0000
africa	0.2168	2.18E-02	4.35E-02	1.0000
gov	0.1016	4.19E-04	1.32E-03	1.0000
landlocked	0.0421	3.11E-03	1.60E-02	1.0000
logpop	0.0395	-6.87E-04	3.68E-03	0.0000
fuel	0.0313	6.55E-05	4.13E-04	1.0000
tradefreedom	0.0171	2.97E-05	2.51E-04	0.9997
invest	0.0129	-3.41E-05	3.36E-04	0.0014
smamerica	0.0117	6.58E-04	6.87E-03	1.0000
spanish	0.0099	5.12E-04	5.93E-03	1.0000
edu3	0.0090	-6.65E-06	8.77E-05	0.1234
dutch	0.0076	-5.14E-04	7.04E-03	0.0000
french	0.0072	3.47E-04	4.95E-03	1.0000
language	0.0068	5.12E-04	7.28E-03	0.9956
goveff	0.0061	-1.27E-04	2.06E-03	0.0114
edu2	0.0056	9.65E-05	1.60E-03	1.0000
fdi	0.0050	-3.90E-05	6.83E-04	0.0000
loggdppc	0.0042	-3.24E-05	9.18E-04	0.2907
exports	0.0041	-4.59E-06	1.32E-04	0.0346
tot	0.0037	5.69E-06	1.19E-04	1.0000
europe	0.0036	-1.04E-04	2.37E-03	0.0012
namerica	0.0036	-1.69E-04	3.63E-03	0.0000
polityIV	0.0032	-4.26E-06	1.78E-04	0.1557
trade	0.0029	-4.31E-07	7.03E-05	0.4231
capital	0.0029	-4.18E-06	1.41E-04	0.0052
loginfl	0.0028	-2.76E-05	9.18E-04	0.0328
island	0.0028	-6.28E-05	1.94E-03	0.0322
imports	0.0027	1.91E-06	1.13E-04	0.7497
volint	0.0027	7.87E-06	3.28E-04	0.8776
edu4	0.0025	-8.88E-07	4.71E-05	0.4542
asia	0.0025	1.73E-05	1.66E-03	0.3866
british	0.0025	-3.75E-05	1.48E-03	0.0023
tariff1	0.0025	-2.11E-06	9.55E-05	0.1650
volinfl	0.0024	2.58E-06	2.67E-04	0.7173
tariff2	0.0024	-2.27E-06	8.59E-05	0.1316
loginfl	0.0021	-3.42E-08	3.79E-06	0.0869
edu1	0.0021	-1.96E-05	6.82E-04	0.0043
interest	0.0020	1.27E-06	4.58E-05	0.9850
portuguese	0.0019	-5.09E-06	2.22E-03	0.1931
opec	0.0018	2.18E-05	4.46E-03	0.5917
tariff3	0.0018	-5.73E-07	5.49E-05	0.2902
commonlaw	0.0018	-2.24E-05	1.22E-03	0.0028