AN ALGORITHM FOR THE AUTOMATIC CONSTRUCTION OF BAYESIAN NETWORKS WITH LIMITED DOMAIN KNOWLEDGE, AS APPLIED TO THE PREDICTION OF ECONOMIC AND DEVELOPMENT INDICATORS OF 248 COUNTRIES AND WORLD REGIONS

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

FERNANDO JAVIER TORRE-MORA

Dr. Yi Shang, Thesis Supervisor

MAY 2016

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

AN ALGORITHM FOR THE AUTOMATIC CONSTRUCTION OF BAYESIAN NETWORKS WITH LIMITED DOMAIN KNOWLEDGE, AS APPLIED TO THE PREDICTION OF ECONOMIC AND DEVELOPMENT INDICATORS OF 248 COUNTRIES AND WORLD REGIONS

presented by Fernando Javier Torre-Mora,

a candidate for the degree of Master of Science,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Yi Shang

Professor Jianlin Cheng

Professor Alina Zare

Dedicated to everyone who kept me sane on this crazy adventure

You’re the best and I love you all

ACKNOWLEDGEMENTS

I would like to thank Dr. Yi Shang for his invaluable guidance and patience, Dr. Cheng for his kindness and flexibility, Dr. Alina Zare for her leadership and comprehension, and Drs. Yunxin Zhao and Toni Kazak for believing in me.

I would also like to thank the 2014 Arthur J. Fulbright Commission for selecting me, among all the Venezuelan applicants, to come to the University of Missouri, the staff at the Institute of International Education, and the staff at the University of Missouri International Center.

A special thank you to Drs. Carolina Martinez, Marlene Goncalvez, and Masun Nabhan for showing me the wonder of these areas of study, and to Drs. Esther Vidal, Janett Mora, and Fernando Chalbaud for your kind support.

TABLE OF CONTENTS

LIST OF FIGURES

The following is a list of the figures contained in this document. Note that figures from the appendices are excluded.

LIST OF TABLES

The following is a list of the tables contained in this document. Note that tables from the appendices are excluded.

ACADEMIC ABSTRACT

Humans have a natural tendency to express knowledge in terms of generalities, instead of individually measurable variables. However, to make any computational sense of a domain, information must be expressed in terms of measurable variables. Current computational methods either require the domain representation to be expressed in terms these variables (which can be hard for a human in many domains), or seek to discover these relationships by assuming the variables are not generalizable (ignoring human knowledge entirely). This project proposes a method to allow representational machine learning methods to be complemented by expert knowledge even when said knowledge is incomplete. The method is implemented as a Bayesian network construction algorithm using the standard error for a least-squares linear regression (STE) on the training data to evaluate each arc—as such, it is a variation of the method proposed by Friedman, Nachman, and Peér in 1999. A proof is included that this variation allows the computation time to be reduced, from factorial order, to quadratic order, with respect to the number of variables. The domain it is tested on is economic forecasting, using models from the literature. In particular, the method's capabilities in reproducing the eminent Smets-Wouters economic model, and in creating a new concrete model from the abstract ideas in the UNESCO world engineering report, are evaluated. The document concludes by studying the differences when comparing the generated models to each other, and to the Smets-Wouters economic model.

Keywords—Bayesian networks; domain representation; domain knowledge; economic forecasting; knowledge discovery; machine learning; model-based agents; least-squares linear regression, socioeconomics.

INTRODUCTION

High dimensionality and missing values plague the majority of data sources today, adding difficulty to their analysis, and limiting their usefulness and predictive value (Tan, Steinbach and Kumar 2005). The curse of dimensionality also comes into play, requiring large numbers of samples for computational application (Theodoridis and Koutroumbas 2008, Duda, Hart and Stork 2001).

Traditionally, this is addressed through feature selection or dimension reduction methods (Theodoridis and Koutroumbas 2008). However, feature selection may discard information that the user is interested in (Duda, Hart and Stork 2001), and dimension reduction involves transforming the features into components that can be hard to interpret (Bishop 2006).

Yet some computational method is needed to handle the high dimensionality of modern data (Davis Kho 2016, Ramezani, et al. 2010), particularly one that can produce a human-interpretable model (Liu, Cocea and Gegov 2016). Although Neural Networks and Support Vector Machines are popular methods for developing such models, their interpretation becomes an exercise in futility (Keller, Liu and Fogel 2015, Russell and Norvig 2010, Liu, Cocea and Gegov 2016).

In recent years, the interpretability of models has become a growing concern (García, et al. 2009) and growing efforts are being made to improve the interpretability of computed models (Casillas, et al. 2013). Interpretability allows not only a human to understand and trust the computer-generated model, but also to gain "meaningful and useful knowledge from data" (Liu, Cocea and Gegov 2016). Interpretability is traditionally ensured by using a rule-based decision tree (García, et al. 2009), however, these can be difficult to generate computationally and create race conditions and other unpredictable behaviors when rules overlap (Millington and Funge 2009).

A Bayesian Network solves these problems; however, the methods that exist to construct Bayesian networks require either complete knowledge of the domain being modeled (Russell and Norvig 2010), genetic algorithms (Shapcott, et al. 1999), or a perfect topological ordering (Chickering and Heckerman 1997). Few methods exist to construct a Bayesian network using purely mathematical relations between the variables (Cooper and Herskovits 1992) and none allow operating with only partial domain knowledge.

This project proposes such an algorithm, with the intent that it may be used in domains where a full understanding of how the domain works has not yet been reached, and thus, generate knowledge. The methodology described herein is actually general enough to be applied to several representational machine learning methods. As proof of concept, the algorithm is implemented to construct Bayesian networks modelling the economies of different countries. The method exploits the intrinsic categorization of variables, the natural relations between them, and the human tendency to express these relations in terms of general categories.

To test our method, we set out to apply it to the problem of predicting economic growth. This problem not only has a large number of variables on which to build on, but it has also become particularly important in the past eight years given the considerable slowdown that has occurred in the global economy. More informed prediction mechanisms would prove invaluable to policy makers and help them make better decisions. The question ultimately boils down to how important each of these factors is.

Data is obtained from The World Bank. This allows several models to be generated (one for each country and region) and compared, exploring the power of this method to adapt to different data.

The next section explains in greater detail the problem of economic growth. Chapter 2 describes previous work, both in computing factors of economic growth, and in developing Bayesian Network construction methodologies. Chapter 3 gives the formal problem formulation. Chapter 4 explains the complexity of the dataset. Chapter 5 goes into our Bayesian Network construction methodology, followed by a discussion of our results in Chapter 6, both as far as computed networks and our evaluation of them. We conclude by evaluating our successes and remaining challenges.

CHAPTERS

1. MOTIVATION

The global economy is in trouble. The increase in the World's Gross Domestic Product (GDP) has been generally falling for nearly a decade (see Figure 1). Governments worldwide and investors have been forced to cut back on spending (UNESCO 2010), reducing the strength of the actors that have traditionally been expected to spur economies (Drahokoupil 2013). A return to the year-to-year growth observed prior to the Great Recession is desirable (the Great Recession can be observed between the years 2007–2009 in Fig.1). This would indicate a return to a global economy capable of withstanding events such as the Asian Financial Crisis (1997–1999 in Fig.1) and the dot-com crash (2001–2003 in Fig.1) without affecting the overall global trend. However, such a strengthening does not seem likely under current conditions. It is only natural for the global question to be how to achieve this.

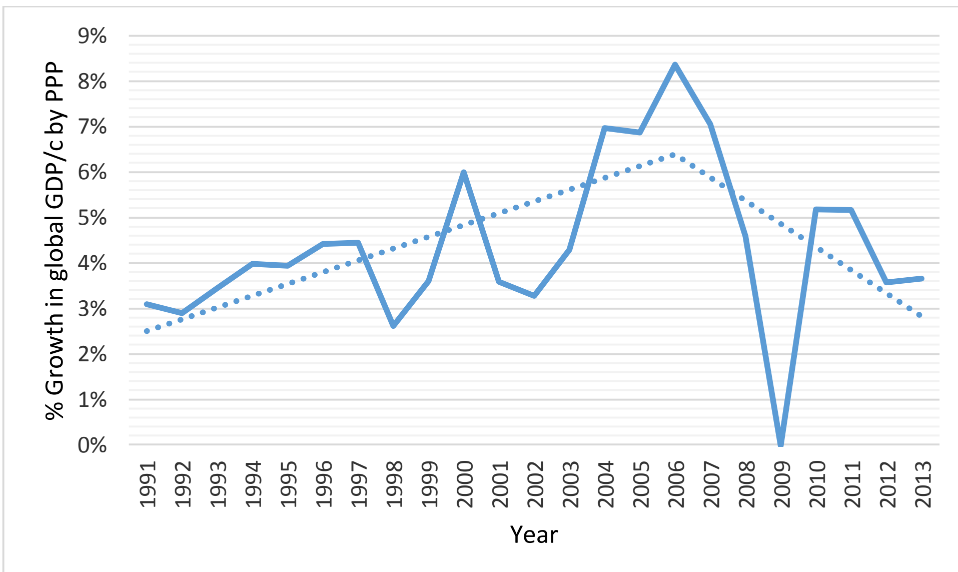


Figure 1: Percentage of growth in worldwide Gross Domestic Product per capita based on Purchasing Power Parity. The percentage ($y_i - y_{i-1}/y_{i-1}$) is shown with a solid line. The trend (as given by an ordinary linear regression on the years covered) is shown with a dotted line.

Recent research has shown that investment in education and technology would be extremely favorable to economic growth. The Director-General of the United Nations Education, Science, and Culture Organization (UNESCO) has stated it thus:

> "The current economic crisis presents challenges and opportunities for engineering. … [T]here are encouraging signs that world leaders recognize the importance of continuing to fund engineering, science and technology[. This investment] may provide a path to economic recovery and sustainable development."
>
> (Bokova 2010)

A country's economic growth has been proven to depend strongly on the number of educated persons that country has in several areas (Jaffe, Rios and Florez 2012, Mora, J.; Torre, F.; Torre, F. 2013, Hausmann, et al. 2011). While the existence of a strong relation cannot be denied, none of these studies have measured the strength of this link compared to other possible factors (Watermayer 2010).

This problem not only has a large number of variables on which to build on, but it has also become particularly important in the past eight years given the considerable slowdown that has occurred in the global economy. Policy-makers have been faced with significant budgetary challenges.

Having had to choose which sector has its funding reduced is non-trivial: policy-makers face deadlock and trade-off to arrive at an effective decision in a timely manner. Often, the negative consequences of these changes are greater than their benefits, such as when education funding is reduced, and the effect has been different in every country.

Worldwide, UNESCO has largely led the fight on preventing education funding from being cut. The 2010 report had the explicit aim to "develop public and policy awareness" on the "social and economic" importance of preparing engineers. A recent article (Finnerty 2014) shows these cuts are still necessary at the state level, and the ultimate decision taken by policy makers have noticeable effects on polls, and hence, on elections.

2. RELATED WORK

2.1. Existing Bayesian Network Construction Methodologies

Procedures for constructing Bayesian networks are scant (Cooper and Herskovits 1992). The most basic method (Russell and Norvig 2010, Chickering and Heckerman 1997) consists of the arrangement of variables in cause and effect ordering and the exploitation of conditional independence assumption such that Chain rule can be applied to form the conditional probability table (Russell and Norvig 2010). This method, which is usually referred as the K2 method due to its use of this measure to evaluate arcs, was first described by Cooper and Herskovits (Chickering and Heckerman 1997, Bouckaert 2008). However, this assumes a perfect topological ordering not only exists, but is known by the user.

Methodologies have been developed to construct Bayesian networks for specific purposes, mainly using genetic algorithms (Helman, et al. 2005, Shapcott, et al. 1999, Chickering and Heckerman 1997); however, due to genetic algorithms' dependence on their initial population (Keller, Liu and Fogel 2015), these algorithms amount to performing the K2 method for several randomly chosen orderings with no controls over how this choice is made.

Comparing a variable against all its possible child nodes using purely their statistical properties has been done previously using the "mutual information" measure of dependence (Friedman, Nachman and Peér 1999). The authors do account for a categorization of the variables; however, they then require these categories to be ordered topologically as in the K2 method above. The authors also point out, that their approach does not seek to optimize any statistical, and indicate that an extension of their method to any graph is necessary.

These methodologies are costly: If we assume a Bayesian model with n source nodes and m sinks, producing any of the models described above takes time $(n+m)!$ The method proposed in this project reduces this time to merely $n \times m$.

2.2. Existing Machine Learning Economic Prediction Methodologies

One of the earliest attempts to apply Machine Learning methods to economic prediction was in 2000. Shaaf (Shaaf 2000) created a neural network with two inputs (difference in treasury bond yields and previous year GDP growth), a single hidden node, and a single output node (estimated GDP growth) for a mean squared error of ten percent. Later that year, Gonzalez (Gonzalez 2000) constructed a neural network using 6 inputs (economic activity growth rate, current employment growth rate, previous employment growth rate, consumer confidence index, real return rate, and government spending), two hidden nodes, and a single output node (estimated real GDP), for a mean squared error of 7.7 percent. It is of note that these results are very hard to replicate for other countries due to the uniqueness of the variables used.

These prediction methods did not gain much popularity—probably because of the interpretability problem inherent to neural networks (Liu, Cocea and Gegov 2016). Instead, the predominant approach was the use of Bayesian Vector Auto-Regression models (Del Negro and Schorfheide 2009). These models behave like Naïve Bayesian Networks, in as much as each variable apart from the target being independent from the others, and the use of missing values producing a response dependent on the frequency of the variable's training values, and the values given for its neighbors (Fernandez Macho, Harvey and Stock 1987). They are also constructed like Bayesian networks, in as much as only a target variable need be selected from the group of input values for the model to be set, after which the conditional probability coefficients can be learned using frequencies. Such models ultimately fell out of favor for Smets and Wouters's model (Lo 2014)

2.2.1. The Smets-Wouters model

Smets and Wouters developed a Bayesian Stochastic Dynamic General Equilibrium model to estimate the behavior of the Euro area as a whole (Smets and Wouters 2002). In traditional Bayesian network construction manner, the model structure was developed manually, while computerized Bayesian methods were used to compute the conditional coefficients of the model. They then adapted the model to the United States (Smets and Wouters 2007)—while other parties adapted the model to England, Canada, and Sweden (Lo 2014)—showing its generalizability. It was adopted by the European Central Bank for use in policy analysis (The European Central Bank n.d.) and is still considered a good benchmark (Del Negro and Schorfheide 2009, King 2012).

The model itself defines economic output (measured as the Gross Domestic Product) as estimable with the sum of all consumption, investment, capital utilization (measured through capital expenditure), aggregate production (measured as "worker service", a function of the number of currently active workers in the economy), and external disturbances (measured using exogenous spending). Current consumption is defined as depending on past consumption, wages (or "labor compensation"), and the interest rate. Investment is defined as depending on the previous year's investment, and "existing capital" (or capital created in the previous year). Capital is define in terms of wages, the difference in capital (or capital accumulation), and the capital-labor ratio (the quotient between the two). Wages are defined as "sticky" and thus depend only on the wages of the previous year and inflation. Finally, capital accumulation is defined to depend only on the capital accumulation that existed previously.

Although the model is presented by Smets and Wouters as a series of equations, it is defined in such a way that it can be represented easily as a Bayesian Network (Figure 2). Variable independence under the Smets and Wouters model is consistent with the functioning of a Bayesian network: for example, if the value for "Consumption" is known, then the value for "Previous year consumption" will not be used to estimate "GDP".

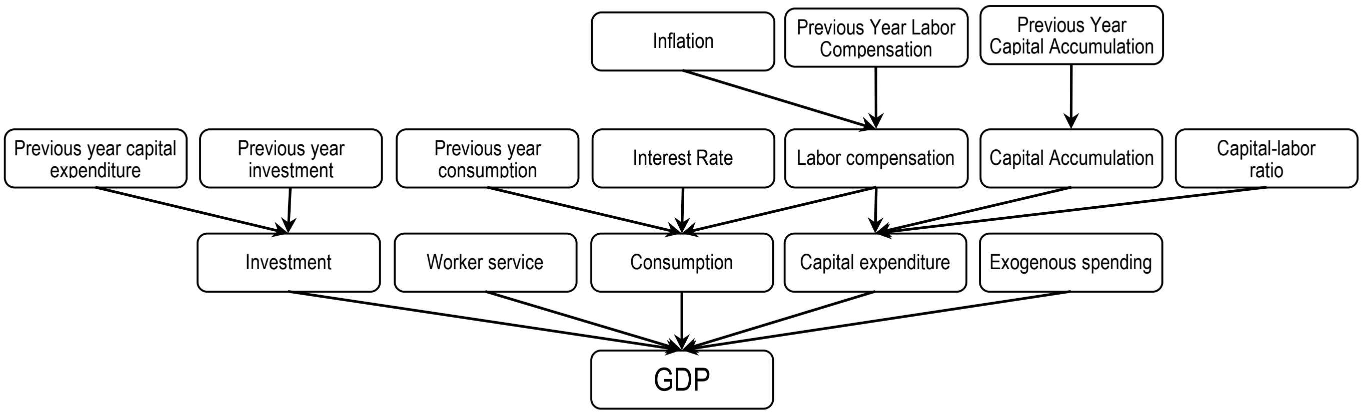


Figure 2: Smets and Wouters economic model (Smets and Wouters 2002, Smets and Wouters 2007). Note that each variable is easily quantifiable.

Note that the model is a Markov process because the estimated GDP growth depends only on values from the current and immediately preceding year (Keller, Liu and Fogel 2015).

When originally presented by Smets and Wouters, the model was evaluated using the Root Mean Square Error (RMSE). Each prediction gives a distribution, of which they take the mean as the predicted value, and compare the difference between it, and the expected value for that year for the absolute error.

Table 1: RMSE performance reported by Smets and Wouters in both papers (Smets and Wouters 2002, Smets and Wouters 2007) on their selected statisticals

	Quarter-to-Quarter estimate for Euro area	*Quarter-to-Quarter estimate for United States*	*Year-to-year estimate for United States*
GDP	0.53	0.63	1.98
Inflation	0.23	0.26	0.5
Capital expenditure	0.14	0.09	0.37
Labor force	0.21	0.46	1.73
Compensation	0.57	0.68	1.82
Consumption	0.58	0.72	1.94
Investment	1.24	1.77	6.88
Overall		-12.48	-3.03

Table 1 offers a comparison of the performances reported by Smets and Wouters. They evaluated their ability to predict seven of their variables. Note that in their original paper (Smets and Wouters 2002) they reported only the quarter-to-quarter accuracy—that is, the accuracy in estimating the next three months, based on the previous three months. They also did not report an overall error.

2.3. The UNESCO model

Smets and Wouters's model, like most economic models of the twenty-first century, is insipient (Lo 2014, Del Negro and Schorfheide 2009). It limits itself to observing how the economy evolves in the short term, rather than giving any indication of how policy may affect it, which leaves it in a poor position to set future policy (Lo 2014) as it was originally intended (The European Central Bank n.d.).

Much has been published linking different economic and education variables to economic growth. In recent years, research has shown that economic development depends strongly on the number of engineers (UNESCO 2010, Mora, J.; Torre, F.; Torre, F. 2013), scientists (Hausmann, et al. 2011, Jaffe, Rios and Florez 2012, Westholm 2010), researchers (Jaffe, Rios and Florez 2012, Westholm 2010), and experts in technology (Jaffe, Rios and Florez 2012, Watermayer 2010). Specifically, economic growth depends on the knowledge gained by these persons that can be used to manufacture goods, perform services, and improve the productivity of existing process. This knowledge is known as "productive knowledge" (Hausmann, et al. 2011, UNESCO 2010) and is usually a subset of the knowledge gained by these persons in their higher education studies, or new knowledge generated by them.

However, the exact weight of productive knowledge in comparison to more traditional factors such as government spending and foreign investment (Drahokoupil 2013) has never been quantitatively assessed (Watermayer 2010). Detriments of favoring any one factor exclusively, are well known, but the strength of the influence is observed through trial and error, if it is assessed at all (ITEP 2013).

The UNESCO World Engineering report (UNESCO 2010) Defines a series of relations which, when taken together, can be used to define a new model:

The number and quality of graduates is strongly affected by economic pressures (Jones 2010), but educated persons are known to be the main drivers of innovation (Jowitt 2010). Innovation, both by individuals and by governments, is a key element and to produce competitively and adequate infrastructure, which must always be done above the existing infrastructure (Miles 2010). Infrastructure cannot grow without investment, and is key to productivity, which in turn is key to the economy (da Silva, et al. 2010). However, even with investment, infrastructure development is dependent on the economic conditions in the country (Stansbury and Stansbury 2010). These relations are mapped out in Figure 3.

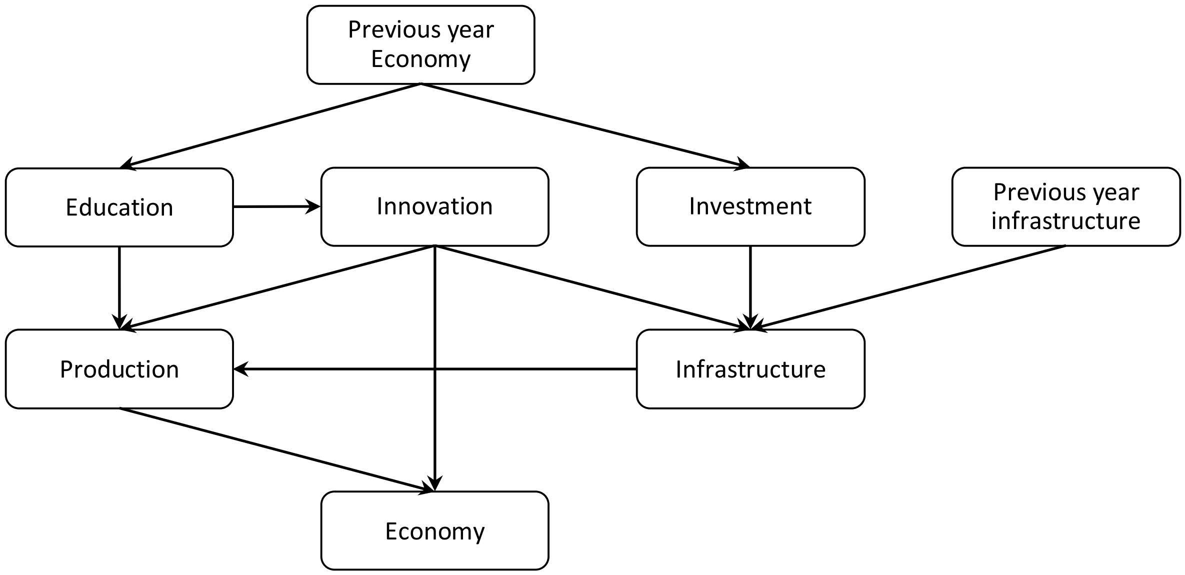


Figure 3: Relations stated by UNESCO shaped into an economic model. Note that everything is stated in terms of general categories

As will be explain in the next chapter, such a model can indeed be used to measure productive knowledge, and thereby the effect of higher education on the economy, by comparing the number of graduates in different areas, and the research they perform. (Jaffe, Rios and Florez 2012, Mora, J.; Torre, F.; Torre, F. 2013).

Note that, once again, the model is a Markov process because the current economy depends only on values from the current and immediately preceding year.

3. GOALS OF THE STUDY

The interpretability of models has become a growing concern (García, et al. 2009) and growing efforts are being made to improve the interpretability of computed models (Casillas, et al. 2013). Interpretability allows not only a human to understand and trust the computer-generated model, but also to gain “meaningful and useful knowledge from data” (Liu, Cocea and Gegov 2016). In this project, Bayesian networks are proposed as a solution to this problem. Bayesian networks not only show where a relationship exists, but also the degree of conditional dependence through its conditional probabilities, allowing estimating unknown values.

Bayesian models are traditionally developed by hand. However, it took Smets and Wouters three years to develop their 2002 Euro area model (The European Central Bank n.d.), and five more merely to adapt it to the United States (Smets and Wouters 2007). An adaptation to the Global economy came in 2004, the United Kingdom in 2005, Canada in 2006, and Sweden in 2007 (Lo 2014). This slowness in adapting and updating models is considered part of the reasons for the Great Recession (Lo 2014).

However, data is being generated at ever increasing rates (Davis Kho 2016), as are categorizations and indeed ontological graphs (Ramezani, et al. 2010). Central banks are trying to accelerate the rate at which they update their predictions without sacrificing interpretability (Barhoumi, et al. 2008). If Bayesian networks are the answer, an automatable construction methodology seems necessary.

Among their difficulties faced by Smets and Wouters were the simple name differences between the variables—they term their variable r_t (Capital expenditure in Table 1 on page 5) “Market investment” in their original work (Smets and Wouters 2002), and “Fed funds” in their U. S. adaptation (Smets and Wouters 2007). This problem is avoided in this project by using variable names from the World Bank, which are standardized. It is also aided by the use of categories, where variables with similar semantic meaning can be grouped together, letting the model automatically pick the best one. Smets and Wouters also go into great length to discuss the viability of keeping each one of the arc relations from the Euro model in the U.S. model; whereas with the proposed methodology, this is taken care of automatically.

Existing algorithmic Bayesian construction methods which ignore the domain knowledge tend to produce counter-intuitive results, hindering their interpretability. For example, preliminary experiments using the K2 method (see section 2.1), as implemented by Bouckaert 2008, on our data, would place variables that reflected the total investment made throughout a year, as causes of the growth generated the previous year. To solve these problems, a Bayesian Network construction methodology is proposed.

However, It seems unlikely—and in fact is strongly discouraged (Migiro 2016)—that a single network can encompass all the countries in the world accurately. Preliminary experiments showed that such an approach is not much better than selecting values at random. Thus, a study is proposed where a network is generated for each country.

The purpose of the study is twofold: The first is, using the Bayesian construction algorithm, to investigate the reproducibility and adaptability of an existing model. The second is to test its capability in deriving a new model and compare it to the existing one.

3.1. Reproducing and expanding upon an expert-derived model

The first goal is to test how close the algorithm can come to a model derived by experts. Because of the difficulty in obtaining the data used in Shaaf's and Gonzalez's models, the project will attempt to replicate Smets and Wouters's model.

To test whether the algorithm can derive it, the variables in the Smets and Wouters's model are divided into five categories or layers. The layers are derived from their "row" in the Bayesian dependency graph (Figure 4), separating variables from the current year from those in the previous year. The algorithm will then receive a graph connecting these layers (Figure 4) and a table indicating the contents of each layer (Table 2) as the domain knowledge and will compute the best way to link the variables using it.

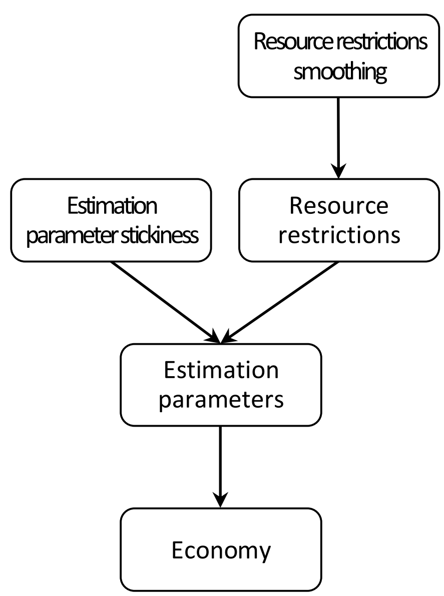


Figure 4: Layer graph representing the domain knowledge shown in Figure 2. Names are those used in Smets and Wouters 2007

Table 2: Categorization of variables in the Smets and Wouters model, providing a mapping from Figure 4 to Figure 2.

Category	*Variables*	*World Bank name*
Resource restrictions	Labor compensation	Compensation of employees (current LCU)
	Interest rate	Lending interest rate (%)
	Capital accumulation	Gross capital formation (annual % growth)
	Capital-labor ratio	[Computed]
Estimation parameters	Consumption	Final consumption expenditure (constant LCU)
	Investment	Portfolio Investment (BoP, current US$)
	Worker service	Wage and salaried workers, total (% of total employed)
	Capital Expenditure	Net capital account (BoP, current US$)
	Exogenous Spending	[Computed]
Economy	GDP	GDP (constant LCU)
Estimation parameter stickiness	Previous year consumption	Final consumption expenditure (constant LCU)
	Previous year investment	Portfolio Investment, net (BoP, current US$)
	Previous year capital expenditure	Net capital account (BoP, current US$)
Resource restrictions smoothing	Inflation	Inflation Consumer prices (annual %)
	Previous Year Labor Compensation	Compensation of employees (current LCU)
	Previous Year Capital Accumulation	Gross capital formation (annual % growth)

Table 2 also shows the World Bank equivalent for each of the variables in the Smets-Woulters model. These were found by hand and allow us to test the model on any country.

Two attributes are marked as computed: Exogenous Spending and the Capital-Labor ratio. These are not reported by the World Bank but are readily computed from other data available. I.E., exogenous spending is defined as

$$\mathrm{ES} = \mathrm{GNE} - (\mathrm{GDP_I} - \mathrm{D_{GDP}})$$

Where ES is Exogenous Spending, GNE is Gross National Expenditure (the sum of all spending done by a country's citizens and corporations at home and abroad), GDP_I is Gross Domestic Product as computed by the sum of all Income ("GDP (constant LCU)" in the World Bank), and D_{GDP} is the Discrepancy between GDP_I and GDP_E (GDP_E is Gross Domestic Product as computed by the sum of all spending within the country's borders; D_{GDP} is "Discrepancy in expenditure estimate of GDP" in the World Bank)

The capital-labor ratio is defined simply as

$$\mathrm{CLR} = \frac{\mathrm{C}}{\mathrm{L}}$$

where CLR is the Capital-Labor Ratio, C is the total Capital expenditure ("Net capital account (BoP, current US$)" in the World Bank), and L is the total labor compensation ("Compensation of employees (current LCU)" in the World Bank)

Of particular interest is, given this Domain Knowledge model, how close can the algorithmically-derived network come. Specifically, it is of interest to compare the networks the algorithm produces given the data form the Euro area and the United States respectively, to Smets and Wouters's original network (Figure 2). This will be termed the reproducibility experiment.

3.2. Deriving a new model from partial domain knowledge

The second goal is to derive a new model. A simplification of the model described in section 2.3 is used for this. This is both to keep the comparison fair (Figure 3 is much more complicated than Figure 4) and because there are no good indicators of infrastructure in the World Bank (data has only been collected on this aspect since 2011). The model to be used is achieved by eliminating the infrastructure categories and connecting the loose edges. The resulting Domain Knowledge graph is shown in Figure 5.

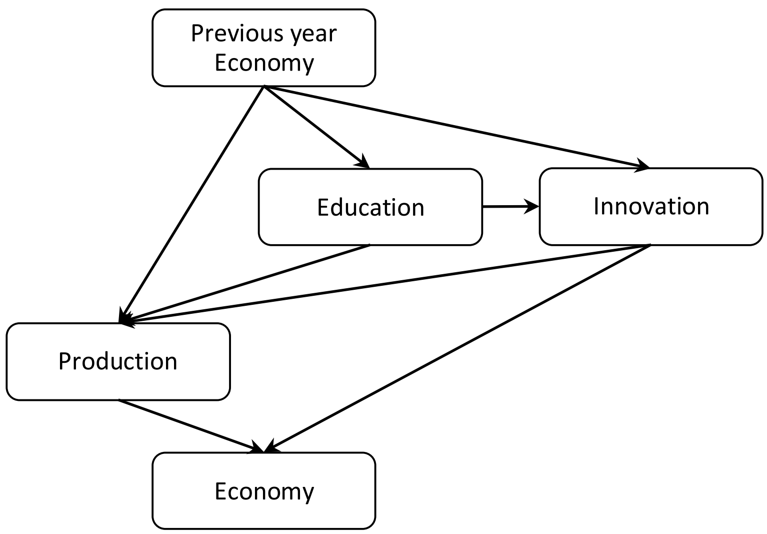


Figure 5: Simplified UNESCO economic model layer graph

As stated previously, each of the nodes in this graph refers to a rather abstract, hard to measure concept, for which reason we must define concrete variables to measure them by.

Fourteen variables were hand-picked from the World Bank Data Bank list of World Development indicators (World Bank 2016). We choose PPP as a suitable measure of the economy independent of inflation, as is traditional in these analyses (Mora, J.; Torre, F.; Torre, F. 2013, Hausmann, et al. 2011, Jaffe, Rios and Florez 2012); Industry value added as a measure of mining, manufacturing, and construction; and Unemployment as a measure of economic activity/inactivity. We use the Unemployment ILO estimate since definitions of employment can vary from country to country whereas the ILO uses a single standard definition for all countries (International Labour Organization n.d.).

Education is quantified in terms of the proportion of the labor force (working age adults) at each educational level. Although the UNESCO engineering report mentions scientists, engineers, and technicians, Mora et al. found that very little improvement is observed when separating into these categories. Labor Force with Primary Education refers to the number of working-age adults that have completed elementary school or its local equivalent, Labor Force with Secondary Education refers to the number of working-age adults that have completed high school or its local equivalent, and Labor Force with Tertiary Education refers to the number of working-age adults that have completed College or its local equivalent (International Labour Organization n.d.).

Innovation is quantified in terms of Journal articles, measuring the amount of research being performed in the region; Trademark applications as a measure of new businesses and products; and Government Expenditure as a measure of government research grants, investments, and incentives in general, covering three of the four Frascati institutions of innovation in the economy (Westholm 2010).

The variables and their categories are summarized in Table 3.

Table 3: Categorization of variables in the UNESCO model. Variable names are those given by the World Bank

Categories	***Variables (World Bank name)***
Economy	GDP per capita, PPP (constant 2011 international $)
	GDP growth (annual %)
Previous year economy	GDP per capita, PPP (constant 2011 international $)
	GDP growth (annual %)
Production	Industry, value added (% of GDP)
	Services, etc., value added (% of GDP)
	Unemployment, total (% of total labor force) (modeled ILO estimate)
Education	Labor force with primary education (% of total)
	Labor force with secondary education (% of total)
	Labor force with tertiary education (% of total)
Innovation	Scientific and technical journal articles
	Trademark applications, total
	General government final consumption expenditure (% of GDP)

4. DATASET

The World Bank open data bank was used as a data source. The World Bank recognizes 217 countries and territories and compiles the data these countries report yearly to the United Nations (World Bank n.d.).

Data is available for more countries for the variables selected for the UNESCO model than for the Smets and Wouters model (Figure 6). This highlights the need for new models, as many of the variables required by Smets and Wouters are simply not reported by the majority of countries.

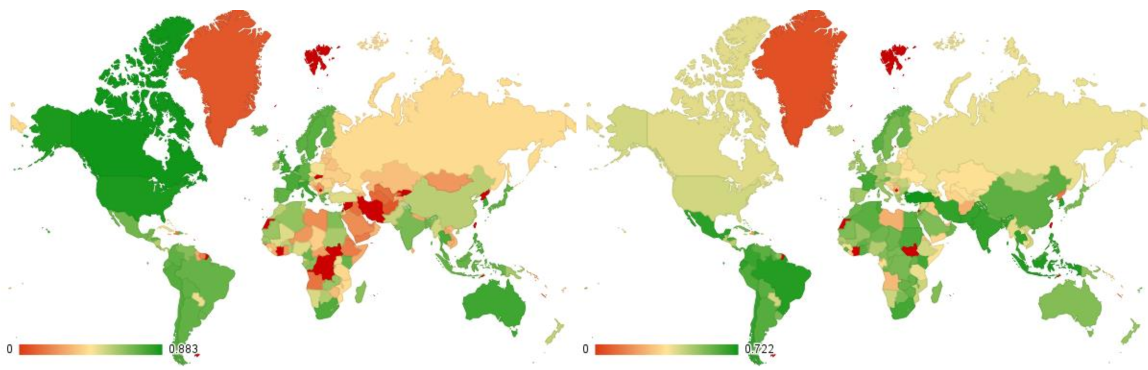


Figure 6: Total proportion of data available for each country, for the variables selected (left) for the Smets and Wouters model and (right) for the UNESCO model. A datum is considered available if its value is not blank. Note that Taiwan, French Guiana, Western Sahara, and Svalbard are not considered countries by The World Bank and are therefore marked as having zero data. To see the proportion of data available for a specific variable, go to Appendix 1 starting on page 64.

Each indicator has data from 1960 to 2015, with some values missing. Data is also available for more years for the variables selected for the UNESCO model (Figure 7), showing that countries gather and report this information more consistently than for the purely economic variables. This is not expected to be a problem, as the Smets and Wouters model for the United States was trained with data from only 34 years: 1966–1972 and 1984–2004 (Smets and Wouters 2007), but suggests that the predictions made by the UNESCO model will be more consistently reliable for more countries (e.g.: the Smets and Wouters model, as trained by Smets and Wouters, probably does not predict very well the effects of the 1979 oil crisis, since its nearest preceding data point to 1979 is 1972, and they report 1978 was the second most deviant year in the twentieth century)

The World Bank additionally defines, for each country, one or more: geographic region, lending group, and economic group (World Bank n.d.). Data from each country tends to behave similar to all its groups. These regions are not shown in Figure 6, but can be looked up in Appendix 2, which starts on page 73.

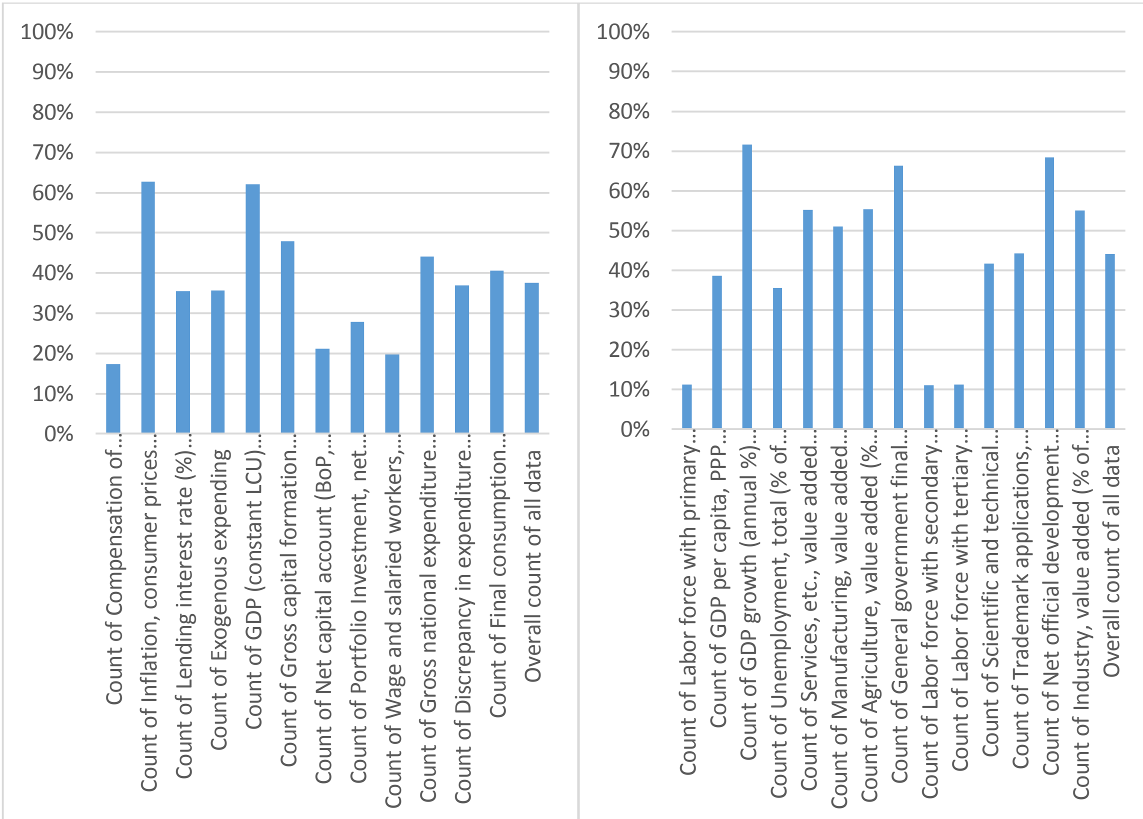


Figure 7: Total proportion of data available for each variable, for the variables selected (left) for the Smets and Wouters model and (right) for the UNESCO model. A datum is considered available if its value is not blank. To see the proportion of data available for a specific country, go to Appendix 2 on page 73.

5. METHODOLOGY

The algorithm is largely performed in three stages: in the first stage, the data is **preprocessed** and split into training and testing. In the second stage, a **constructor** takes the domain knowledge model, and the training data: the training data is used to set the weights of the domain knowledge, and translated into an *untrained* Bayesian network. In the third stage, the generated network is given to an **evaluator**, which trains the network with the training data, tests it with the testing data, and produces the appropriate accuracy measures. This process is summarized in Figure 8.

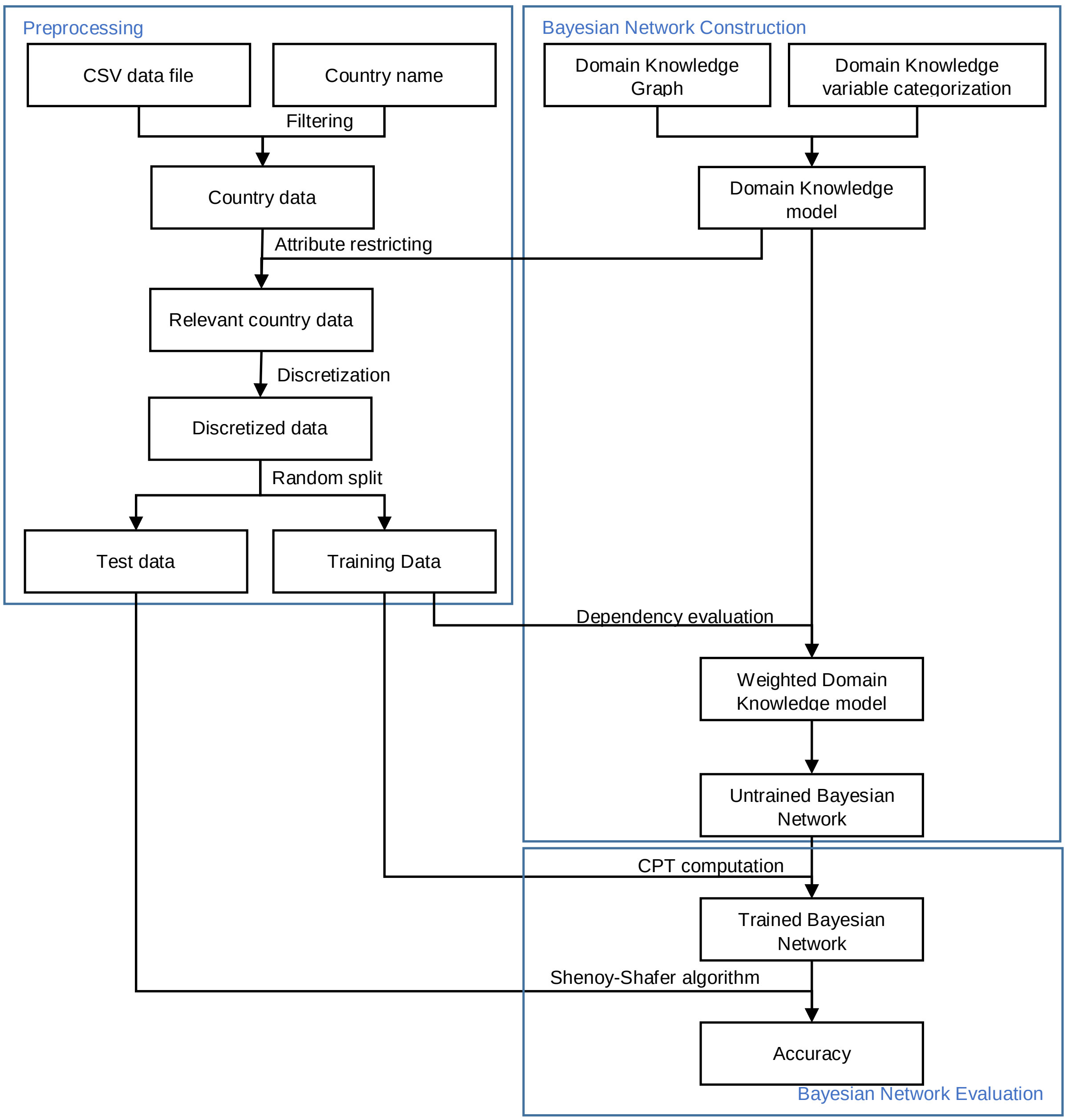


Figure 8: Data pipeline for a single cross-validation fold

This chapter examines each of these, in turn, as well as how they can be enhanced for better performance, then describes the experiments that will be performed to evaluate them.

5.1. Preprocessing

The data was downloaded from the World Bank Open Data Bank (World Bank 2016). All of the variables in Table 2 and Table 3 were selected, as well as all 248 countries and all 65 years. The data was reformatted so that each row was identified by a country or region name, and a year, and each column was a variable.

Because of this, it is necessary to select just the relevant rows and columns before being able to process the data. The appropriate rows are given by a parameter, specifying the country name, for which the "country name" column is then reviewed for matches. The appropriate columns are taken from the domain model being used and compared against the title row of every column. For this reason, when specifying domain models, the name of a variable must be an exact match to its name in the data.

The data from the World Bank does not readily include data from the previous year. This is taken by the program as part of preprocessing. The program also takes as parameter a key word which will be sought in all the variable names of the domain model as a prefix indicating that the variable should be taken from another year. This parameter is called the **shift-by prefix**. Shifting is achieved by creating copy of the variable array and inserting a missing value in the first position, shifting all other values down (and losing the last value). This allows populating these variables with the values from the immediately preceding year with little effort. The program is also capable of shifting by more than one year by specifying different **shift-by amounts**, designed for multi-step Markov processes, but in our experiments we always leave it at 1.

5.1.1. Data discretization

Because Bayesian networks tend to perform better on discretized data (Russell and Norvig 2010), all variables are discretized using the Weka equal-frequency discretizer (Trigg and Frank 2015). This means that each attribute's values are divided into ranges such that each range contains about the same number of values. This is done in the interest of fairer results, as well as to improve training conditions (no single discretized value becomes under-sampled).

In preliminary tests, using 3 bins was found to give good accuracies, and produce easy to read ranges (good performance and interpretability).

5.1.2. Training/testing data split

A 0.85-0.15 separation of the data is performed; that is, 85 percent of the data is set for training and 15 percent for testing (see Table 4 for a full list of parameters). When using a country’s 65 data points, this means that 55 data points are used for training, and 10 for testing.

Because data is indexed by year, to split the data, picking 55 random points amounts to picking 55 random years. This is achieved by reordering the years randomly, and taking the first 55 for training.

Table 4: Summary of parameters used for Preprocessing

Input Parameter	*Value(s) used*
Data file	Data extracted from World Development Indicators.csv (World Bank 2016)
Domain knowledge model	Smets-Woulters model; UNESCO model
Number of discretization bins	3
Discretization type	Equal frequency
Training set size	85%
Shift prefix	“Previous”
Shift amount	1

5.2. Bayesian Network Construction

The Bayesian Network constructor receives the domain knowledge model, in the form of a graph—showing the relations between the categories—and a table—showing the variables contained in its categories—(examples can be seen in Chapter 3). It also receives the training data, a **threshold dependency difference**, and a **minimum forward dependency**, which will be used to evaluate arcs.

The constructor iterates over every edge in the domain knowledge graph. Let the node at which the edge starts be called the “head category” and the node at which the edge ends be called the “tail category”. The constructor can then look up the corresponding variables for both categories in the domain knowledge categorization table, and calculate the strength of the dependency between them. This, in effect, is a dense subgraph, where there is an arc from each “head” variable to all “tail” variables, and the weight of the arcs is the dependency score. The subgraph is restricted in such a way as to follow the domain knowledge’s indication of which variables should have parents and which variables should have children (i.e. no arcs are added from the “tail” variables to the “head” variables).

The subgraph can be stored using any valid graph representation. In the implementation used for this project, a simplified adjacency matrix was used, where the “head” variables were columns and the “tail” variables were rows. That is, omitting the source vertices from the rows and the sink nodes from the columns.

The method for arc evaluation is based on the standard error for a least-squares linear regression, or STE (Microsoft 2007). This metric is consistent with commonly used statisticals (Friedman, Nachman and Peér 1999, Zady 2000) and has the additional advantage of allowing us to get an indication of causation. This, combined with minimal domain knowledge, allows defining an unambiguous, valid Bayesian network.

5.2.1. Arc Evaluation

STE is used to calculate dependence (Mora, J.; Torre, F.; Torre, F. 2013, Microsoft 2007). It is important to note that the methodology is not tied to this statistical (Friedman et al. suggests correlation or mutual information for this purpose; however, STE is known to be consistent with both of these statisticals (Zady 2000)). STE is used because it is a non-symmetric (STE(Y,X) $\neq$ STE(X,Y)) indication of which variable is the dependent variable and which one is the independent variable. For clarity, STE of two variables X and Y will be denoted as $\mathrm{STE}(X \rightarrow Y)$, with the arrow indicating which variable is presumed independent and which one dependent.

STE is defined in such a way that a small STE($X \rightarrow Y$) is associated with a strong causative relation where Y depends on X. It is important to note that STE is not a decisive measure of causation; it simply indicates which variable should be the function parameter and which one should be the function result, assuming the function is linear—which is a reasonable assumption, and one that's traditionally made, in the field of economics (Fernandez Macho, Harvey and Stock 1987, Lo 2014, Del Negro and Schorfheide 2009).

The formula for STE is shown below with Y and X being vectors of values that have a length n, and with $\bar{Y}$ and $\bar{X}$ being their respective sample means.

$$\mathrm{STE}(X \rightarrow Y) = \sqrt{\frac{1}{n-2}\left(\sum_{y \in Y}(y-\bar{Y})^2 - \frac{\left(\sum_{x \in X, y \in Y}(x-\bar{X})(y-\bar{Y})\right)^2}{\sum_{x \in X}(x-\bar{X})^2}\right)}$$

To make the result easier to interpret, the **dependency** equation from Mora et al. is used. Using this equation means higher values are better, and that the measure is normalized into the [0, 1] range; STE is always in the $(-\bar{Y}, \bar{Y})$ range (Mora, J.; Torre, F.; Torre, F. 2013). Note that vertical bars denote absolute value (positive root).

$$\mathrm{dependency}(X \rightarrow Y) = 1 - |\mathrm{STE}(X \rightarrow Y)|/\bar{Y}$$

Because, again, this is not symmetric, we shall call **forward dependency of X→Y** $\mathrm{dependency}(X \rightarrow Y) = 1 - |\mathrm{STE}(X \rightarrow Y)|/\bar{Y}$ and **backward dependency of X→Y** $\mathrm{dependency}(Y \rightarrow X) = 1 - |\mathrm{STE}(Y \rightarrow X)|/\bar{X}$.

Consider the case when the forward dependency is simply too small to imply a causal relationship (irrespective of the value of the backward dependency). For this purpose, use the **minimum forward dependency**, a user-defined number which, thanks to the dependency formula converting the STE result to the [0,1] range, is between zero and one.

5.2.2. Network production

To decide whether to add an arc to the Bayesian network, further define the **dependency difference** as $\text{dependency}(X \to Y) - \text{dependency}(Y \to X)$. This number should be greater than a **threshold** if Y truly does depend on X. This threshold is a user-defined parameter.

Note that if the dependency difference is less than the threshold, or indeed $\text{dependency}(X \to Y) < \text{dependency}(Y \to X)$ rather than concluding Y is the cause and X is the effect, it is discarded entirely. This is because the comparison is made following the links in the Domain Knowledge Model, and adding such an arc would contradict the domain knowledge (cf. counter-intuitiveness discussion at the start of chapter 3).

In our preliminary tests, a minimum value of 0.5, and a threshold value of 0.03 (slightly greater than zero) were found to produce an intuitive number of valid arcs. This is because the data is prone to errors and outliers: for example, the Democratic Republic of the Congo reports having paid only 0.94 Congolese Francs total of worker compensation in 1990, even though for most other years they report numbers in the order of billions. A summary of the parameters used is shown in Table 5 (where "plot mode" refers to how the nodes are positioned for viewing).

Table 5: Summary of parameters used for the Bayesian Constructor

Input Parameter	*Value(s) used*
Input data	Previously-split training data
Plot mode	Layered
Domain knowledge model	Smets-Woulters model; UNESCO model
Minimum forward dependency	0.5
Threshold dependency difference	0.03

The result is a Bayesian network graph. Bayesian networks are encoded into the XLM-BIFF format—which has the de-facto standard (Gagliardi Cozman 2001, Bouckaert 2015, Cascio and the Automated Reasoning Group 2010)—using the *SamIam* belief network library (Cascio and the Automated Reasoning Group 2010). For more information on how the library is used, see Appendix 3: Class Diagrams on page 86.

5.2.3. Complexity analysis

The runtime of the construction algorithm depends strongly on the Domain Knowledge Model and how many variables it receives (n).

In the best case, each category will have exactly one variable, which would imply the Bayesian network structure is already known, and merely needs to be simplified. In this case, the algorithm performs $2m$ operations, where m is the number of arcs in the Domain Knowledge Model. (The two operations performed for each arc are dependency$(X \rightarrow Y)$ and dependency$(Y \rightarrow X)$.) Since a Bayesian network must be a directed acyclic graph, this case may have m being anywhere between $n-1$ (Markov chain) and $n(n-1)/2$ (transitive closure of a fully reachable graph), where n is the number of variables. Therefore, the algorithm is $\Omega(n)$ and $O(n^2)$ in the best case. This is comparable to the best case in Friedman et al. where each node has one or two candidate parents.

In the worst case, each category has the same number of variables: n/c where c is the number of categories and n is a multiple of c such that $n \geq 2c$. To evaluate each arc, the members of each category in the arc's source have to be compared with the member of each category in the arc's sink, each comparison of which requires two operations, or $2(n/c)^2$ per arc between categories for a total of $2m(n/c)^2$. Since, again, there may be anywhere between $m = n-1$ and $m = n(n-1)/2$ arcs between categories, the algorithm is $\Omega(n^2)$ and $O(n^4)$ in the worst case. This is much better than the worst case in Friedman et al. where all other nodes are candidate parents, leading to $O(n!)$.

It should be noted that, given the nature of Bayesian networks as inference engines, the worst case is highly unlikely to be encountered in practice (unlike in Friedman et al.). In general usage with other datasets and domains, it is more likely that there will be a category with much less variables than the others, since there is always a small group of target variables (usually one, which is Friedman's worst case). For this reason we can assume an average order of n^2.

5.2.4. Missing Value Treatment

Most Bayesian construction algorithms based on data require the training data to contain absolutely no missing values, for which reason the data is culled, leaving only the data points where all attributes are known (Russell and Norvig 2010, Shapcott, et al. 1999, Friedman, Nachman and Peér 1999). However, because the proposed method evaluates every pair of nodes independently, missing values only need to be removed from the pair.

That is, when evaluating an arc, only two vectors are necessary: one with the values of the independent variable (X) and one with the values of the dependent variable (Y). An element cannot be added to the summation if the value in either vector is missing (e.g.: $y = Y_i$ cannot be added to $\sum_{y \in Y}(y - \bar{Y})^2$ if X_i's value is missing). However, this is irrespective on whether the value i in any other attribute is missing. This means that, for any entry i, the value for an attribute will be ignored only for the arcs connecting it to an attribute whose value in i is missing, and will be considered for all others that are not missing.

5.3. Bayesian Network Evaluation

The main means used to test the predictive accuracy of the Bayesian networks produced is cross-validation.

5.3.1. Cross-validation

Several random splits are performed on the data, as described in section 5.1.2, each of which is then evaluated for accuracy. The average accuracy of all the splits is reported

In preliminary experiments, we found that 20 splits produced the most stable results. A summary of the parameters used is shown in Table 6.

Table 6: Summary of parameters used for the Bayesian Evaluator

Input Parameter	*Value(s) used*
Input data	Discretized data for the target country
Domain knowledge model	Smets-Woulters model; UNESCO model
Group-by column	Country code
Discretization type	Equal frequency
Training set size	85%
Number of cross-validation folds	20

5.4. Evaluation Experiments

The following experiments are proposed, additional to those from Chapter 3, to be applied to all resulting Bayesian networks.

5.4.1. Reconstruction power

Both Bayesian network and economic models are usually evaluated by means of its predictive power: their ability to correctly indicate the value of a datum they have not seen before. This is done by taking the data from the testing set, and removing one of the attributes.

The Bayesian network is instantiated with the values of one of the years in the training set, for all of the attributes, except the one it is being tested for. The probabilities for each of the discrete values of that attribute are then computed, and the value with the highest probability is taken as the predicted value. This value is then compared with the removed attributes actual value. If matched, the prediction is counted as correct. If there is no value, the datum is omitted in computing the accuracy for this attribute. The accuracy is thus:

$$\text{accuracy}(BN, T, E) = \frac{|\{BN(T_i) = E_i\}|}{|\{E_i \neq \text{null}\}|}$$

where BN is the Bayesian network, T is the test data with the target attribute removed, E is the expected values for that attribute in the test data, vertical bars denote cardinality, and “null” denotes a value is missing.

If all values for that variable are missing in the test set, an accuracy is simply not reported for that split.

Note that, while indicative if the model’s predictive power, a random split could result in data from future years being used, and thus, not a true prediction of the future. This experiment is therefore termed “reconstructive power”, since it is better suited as an indication to reconstruct historical missing values. Given the large number of missing values in the data, this application is no less important than true prediction of the future.

5.4.2. Matching network experiment

It is inevitable that two countries will produce the same untrained Bayesian network. Such countries. Such countries are expected to have similarly functioning economies due to geographical or cultural reasons. These similarities seem worthy of analysis.

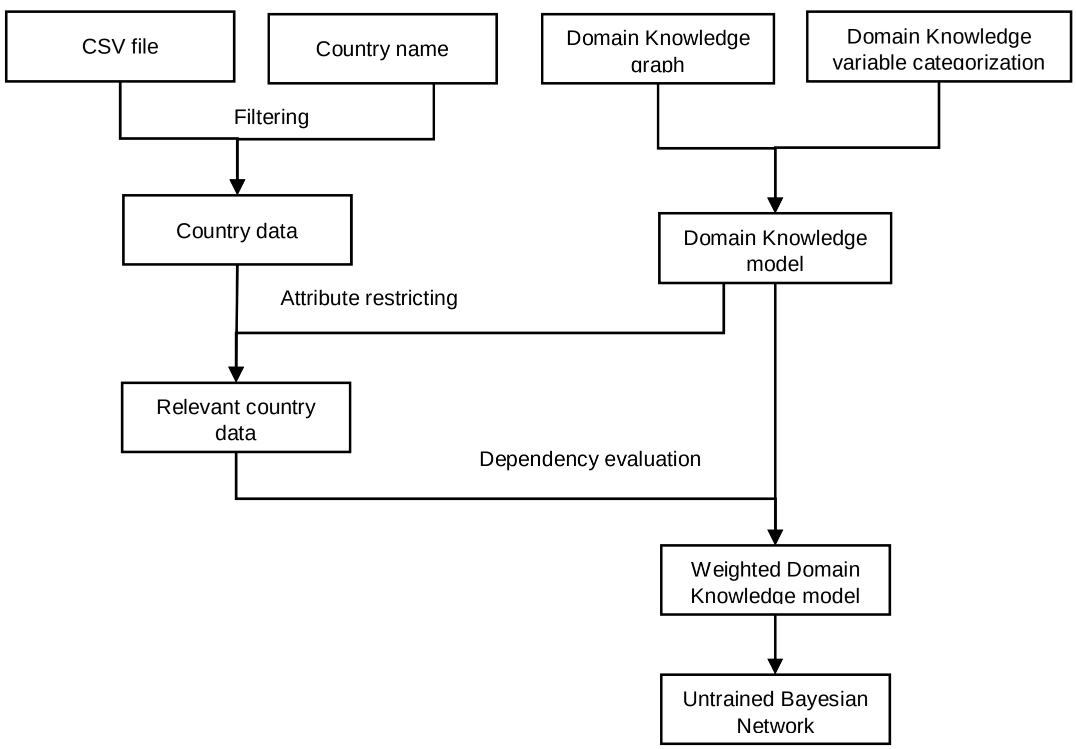


Figure 9: Data pipeline for each country for the Matching network experiment

To detect such matches, a hash table is defined where the keys are the Bayesian networks themselves, and the values are the list of countries with those networks. The hashing defined for SamIam's Bayesian networks is used (Cascio and the Automated Reasoning Group 2010).

Such an experiment is a degenerate clustering since only exact equality is allowed in the "clusters". Because the Matching network experiment does not require training the networks to analyze its results, and is, in fact, an unsupervised learning experiment, it has a slightly different dataflow (Figure 9).

5.4.3. Benchmarking experiment

Finally, we will measure the amount of time each of the following takes for each data split:

- Bayesian network construction time
- Bayesian network training time
- Total processing time

The average will be reported for all the splits performed on the data of the same country. Correlations will be analyzed between these three times, the number of arcs generated and the amount of data available for that country.

6. RESULTS

In this chapter, the networks generated by each model are analyzed, and then compared. The discussion for each model, and later comparison, begins with a discussion on the number of arcs, since this is indicative of the complexity of the Bayesian network generated. Then, each experiment is described, in turn.

For the Smets and Wouters Domain Knowledge model, the first experiment is the reproducibility experiment, comparing its generated results to the original Smets and Wouters model, as described in section 3.1. For both models, the Reconstruction experiment is analyzed next, discussing the average accuracies obtained by each variable among all countries, each country reporting the average accuracy from cross validation. Specifically, the distribution of these accuracies is analyzed and explained. Special attention is paid to the variable with the highest accuracy and the variable with the lowest accuracy. (The individual accuracies for each country are reported in Appendix 6 starting on page 124.) Then the matching network experiment is analyzed, describing which networks were found to be shared. Finally, the Benchmarking experiment is discussed, showing the measured execution times and explaining how they are distributed

Bayesian networks, when they appear, are presented in a layered arrangement, where each row corresponds to one of the layers, or categories, from the domain knowledge model. These graphs are generated by the program in XML BIF format and are captured graphically using the SamIam Bayesian Network viewer. Smets and Wouters's original model is presented in this format in Figure 10. Note that because node placement is automatic, arc overlap is inevitable.

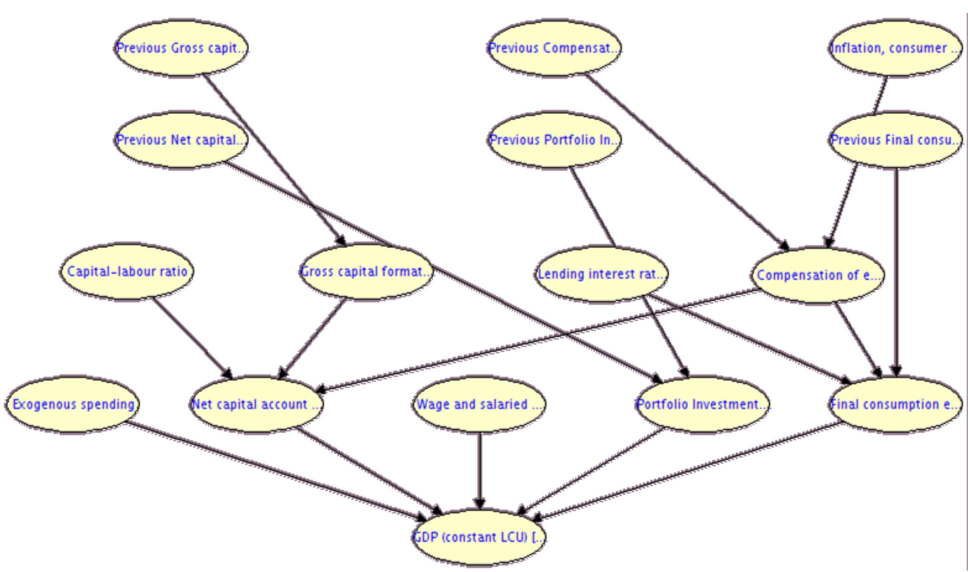


Figure 10: Smets and Wouters original model as a Bayesian network produced by the proposed program. Note that this graph is isomorphic to the one in Figure 2 after mapping the node names to World Bank names using Table 2. Note also that each row of nodes corresponds to a category from Figure 4 after mapping using Table 2.

6.1. Networks generated using Smets and Wouters domain knowledge

The number of arcs generated seems to vary widely for each country. One intuitive explanation is that the number of arcs depends on the amount of data available. And at first glance, with a correlation coefficient of 0.5312, it does seem to have an influence. However, observe in Figure 11 that a function of the data availability would not approximate the number of arcs generated—at best it could estimate the maximum number of arcs. This is confirmed by calculating dependency(availability → arcs), which gives a low 0.2515 (dependency is defined in section 5.2.1 on page 18). This suggests the Bayesian networks truly are capturing the relations between the variables regardless of how much data is missing, to the highest degree the data allows.

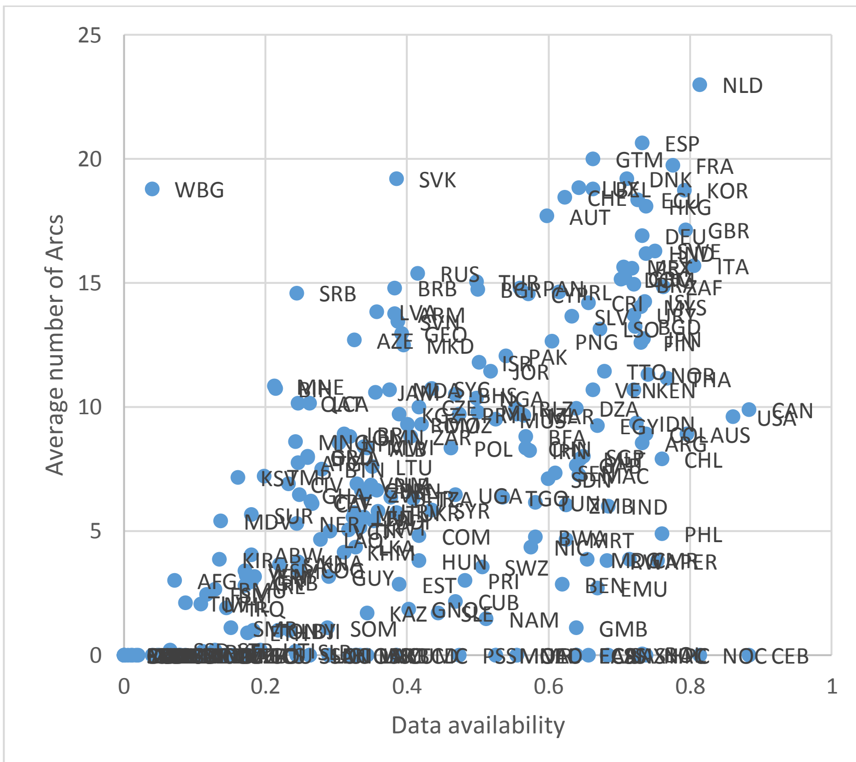


Figure 11: Scatter plot showing the correlation between the amount of data available, and the number of arcs generated, on average, with such data when using the Smets and Wouters Domain Knowledge model. The country code is highlighted for each point. Notable outlier WBG refers to The West Bank and Gaza, for which almost all attributes are available but only after it became a self-governing territory in 1993 (Israel, which collected the data for that area before that point, has no interest in reporting it to the World Bank)

6.1.1. Reproducibility experiment

Smets and Wouters's economic model could not be reproduced. When attempting to reproduce the model for the United States, only four of the fifteen original arcs were found. When attempting to reproduce the model for the Euro area, only two of the fifteen original arcs were found. The arcs found are highlighted in Table 7.

Table 7: Arcs from the Smets and Wouters original model added by the Bayesian construction algorithm

Arc Source	*Arc Destination*	*Constructed USA network*	*Present in EMU*
Previous net capital account	Portfolio investment net	Present	Present
Previous Portfolio investment net	Portfolio investment net	Absent	Present
Compensation of employees	Final consumption expenditure	Present	Absent
Exogenous spending	GDP	Present	Absent
Net Capital account	GDP	Present	Absent

When observing the generated networks, it is interesting to note that some related relations were detected. In the case of the United States, for example, although "Compensation of Employees" was not linked to "Net capital account", as it is in the original model from Smets and Wouters, "Previous compensation" was linked to "Gross Capital Formation". "Inflation" was not linked to "Compensation of Employees", but it was linked to "Lending Interest Rate", both of which affect "Final Consumption" in the original model. "Wage and Salaried Workers", taken to be completely random by Smets and Wouters, has a host of nodes to condition it in the derived model, suggesting that, at least for this node, the derived model is more robust. But perhaps most interestingly, much less arcs were drawn to GDP, indicating that the dependency relation between the estimation parameter variables was not as strong as Smets and Wouters intuited, and giving support to the need for a new economic model, such as the UNESCO model. The generated network for the United State can be seen in Figure 12.

Smets and Wouters only report the accuracy of five of their variables (see Table 1 on page 5). When testing the accuracy, for the other variables, accuracies were found to be much lower. The generated model is able to come close (less than 10 percent difference) or improve on the accuracies for these variables, suggesting that it is more comprehensive than the one originally designed by Smets and Wouters and thus better overall (Table 8).

In the case of the Euro Area, much fewer relations could be detected, probably on account of the lower data availability (see Appendix 1.1 on page 64). As before, conditional dependencies were found for "Wage and Salaried workers", taken to be completely random by Smets and Wouters, but perhaps most interesting is the conditional dependencies found for "Portfolio investment": not only were the dependencies proposed by Smets and Wouters discovered, an additional arc was found linking "Lending interest rate" to it, and "inflation" to "Lending interest rate". These are well-understood dependencies and widely-used in setting policy (Smets and Wouters 2002), which suggest this method may be uniquely suited to identifying relationships human network designers leave out—possibly because they take them for granted. The generated network for the Euro Area can be seen in Figure 13. The tested accuracy is reported in Table 9. Note that only five variables had enough data for an accuracy to be returned (see section 5.4). The low data availability (Appendix 2.1) is normal for regional aggregates (World Bank 2015).

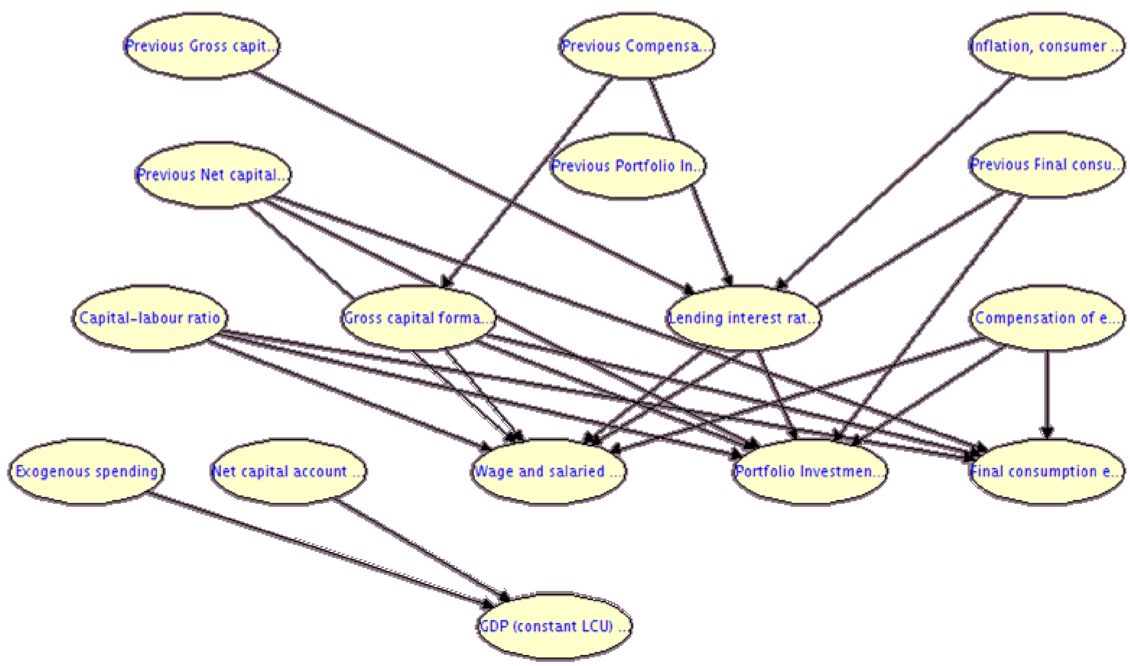


Figure 12: Generated network for the United States using Smets and Wouters Domain Knowledge model

Table 8: Comparison of accuracies obtained by Smets and Wouters original network (Figure 10), and the generated network (Figure 12) on the United States discretized data. Variables originally reported by Smets and Wouters (Table 1) are marked with asterisks. Notice that the unreported variables have the lowest accuracy for the original network.

Variable	*Accuracy of original network*	*Accuracy of generated network*	*Accuracy gained/lost*
Capital-labor ratio	0.8738	0.8158	-6.63%
*Compensation of employees **	0.6222	0.6458	+3.79%
Exogenous spending	0.7509	0.6973	-7.13%
*GDP **	0.9881	0.8330	-15.69%
Gross capital formation (current LCU)	0.9940	0.7107	-28.50%
*Inflation, consumer prices **	0.9082	0.5125	-43.57%
Lending interest rate	0.4405	0.7250	+64.59%
Net capital account	0.7381	0.6025	-18.37%
*Portfolio Investment, net **	0.8173	0.7669	-6.16%
*Final consumption expenditure **	0.9082	0.6875	-24.30%
Wage and salaried workers, total	0.5073	0.7039	+38.77%
Average	*0.7546*	*0.6856*	*-9.14%*

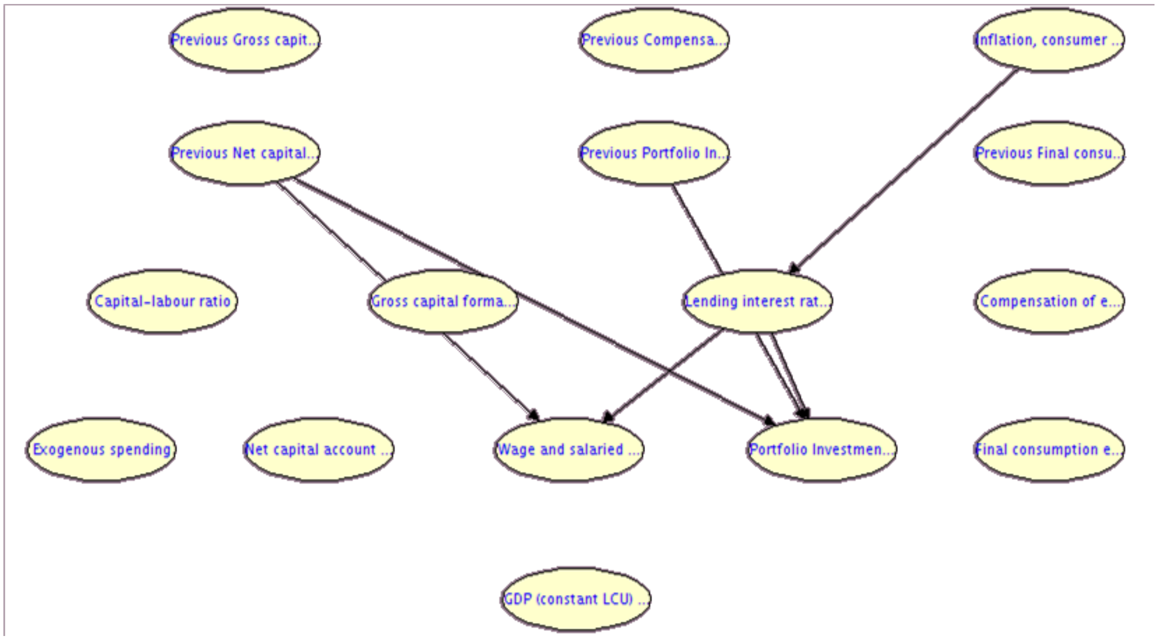


Figure 13: Generated network for the Euro Area using Smets and Wouters Domain Knowledge model

Table 9: Comparison of accuracies obtained by Smets and Wouters original network (Figure 10), and the generated network (Figure 13) on the Euro Area discretized data. Variables originally reported by Smets and Wouters (Table 1) are marked with asterisks.

Variable	*Accuracy of original network*	*Accuracy of generated network*	*Accuracy gained/lost*
Wage and salaried workers, total	0.3206	0.2875	-10.33%
Net capital account	0.3930	0.5167	+31.47%
Lending interest rate	0.4657	0.0139	-97.02%
*Inflation, consumer prices **	0.6607	0.6250	-5.41%
Portfolio Investment, net	0.4403	0.5625	+27.74%
Average	*0.4561*	*0.4011*	*-12.05%*

6.1.2. Reconstruction experiment

The average accuracy for all variables in all countries was 0.5705, which represents a 71.64 percent over choosing a discretized bin at random (accuracy of one-third). For most variables, more countries had accuracies above the mean than below it; i.e.: the population of accuracies was denser in the higher quartiles—this can be observed in Table 10. This distribution is maintained both within geographic regions (Figure 14) and economic groups (Figure 15).

Table 10: Interquartile ranges for the distribution among all countries of the average test accuracies (average among the 20 random splits)

Variable name	*Avg*	*Min*	*Q1*	*Med*	*Q3*	*Max*
Wage and salaried workers, total	0.4707	0.0000	0.3383	0.5000	0.6567	1.0000
Lending interest rate	0.5454	0.0139	0.3577	0.5567	0.7330	0.9775
Portfolio Investment, net	0.6250	0.0000	0.5101	0.6574	0.7931	0.9845
Net capital account	0.6666	0.0294	0.5380	0.6943	0.8581	0.9868
Gross capital formation	0.7272	0.0417	0.6510	0.7616	0.8580	0.9804
GDP	0.6157	0.1000	0.4842	0.6317	0.7768	0.9700
Exogenous spending	0.5741	0.0000	0.4431	0.6182	0.7190	0.9781
Compensation of employees	0.6181	0.0000	0.5046	0.6671	0.7988	0.9833
Inflation, consumer prices	0.5024	0.0000	0.4313	0.5181	0.6025	1.0000
Capital-labor ratio	0.5772	0.0000	0.4042	0.6042	0.8108	0.9868
Previous Final consumption expenditure	0.5101	0.0000	0.4287	0.5401	0.6182	0.9737
Previous Portfolio Investment, net	0.6440	0.0000	0.5206	0.7008	0.8321	1.0000
Previous Net capital account	0.6494	0.0417	0.4838	0.6957	0.8542	1.0000
Previous Compensation of employees	0.6007	0.0000	0.4961	0.6611	0.7941	0.9825
Previous Gross capital formation	0.6936	0.0000	0.5842	0.7348	0.8518	0.9692
Average for all variables	*0.5705*	*0.0000*	*0.5113*	*0.5990*	*0.6555*	*0.8450*

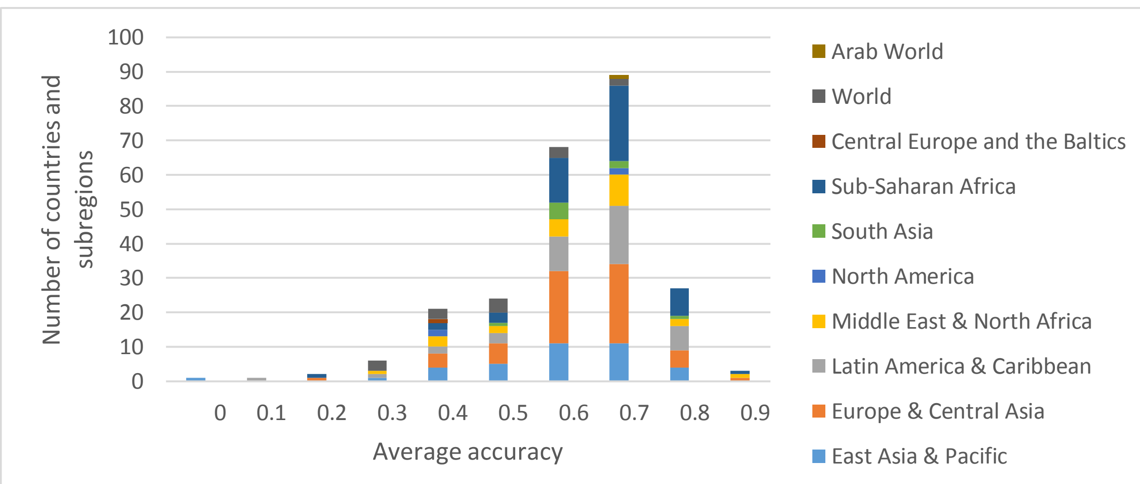


Figure 14: Histogram of average accuracies for all variables for the Smets and Wouters Domain Knowledge Model Reconstruction Experiment, with each country's main geographic region highlighted. The country with zero accuracy is American Samoa, for which the only attribute available is the number of wage and salaried workers.

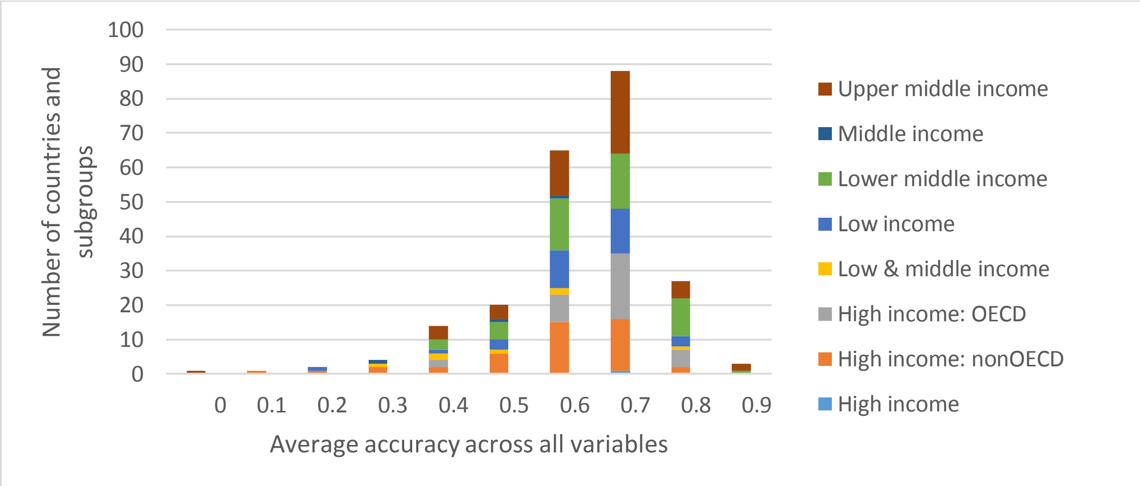


Figure 15: Histogram of average accuracies for all variables for the Smets and Wouters Domain Knowledge Model Reconstruction Experiment, with each country's main economic group highlighted.

This bell shape is exactly what should be expected for a model that generalizes well. The different accuracies can be partially attributed to the lack of data: the correlation coefficient between the average accuracies and the available data is of 0.3952 (see Figure 16). The remaining differences stem from how well the model itself adapts. This suggests that the networks have, on average, good tolerance to missing data.

The variable with the highest average accuracy was "Gross Capital Formation" with 0.7272, representing an improvement of 118 percent over choosing the discretized bin at random. Of note is that the accuracy is comparatively lower for The United States, Canada, and Europe (Figure 17). This makes sense when considering that the reconstruction experiment was not able to capture the nuances for this variable, which the original model (specifically designed for these countries) is able to obtain a very high accuracy for, suggesting that capital formation does indeed function differently in these countries.

The variable with the lowest average accuracy was "Wage and salaried workers" with 0.4707, representing a 41 percent improvement over choosing the discretized bin at random. This makes sense when considering that this variable was the one with the second-lowest amount of data available (see Figure 7 in Chapter 4; page 14). Further, this variable is treated as completely random by Smets and Wouters—with only one outgoing connection to GDP—and the improvement their network gains by doing so with respects to the result presented here is only of three percentage points.

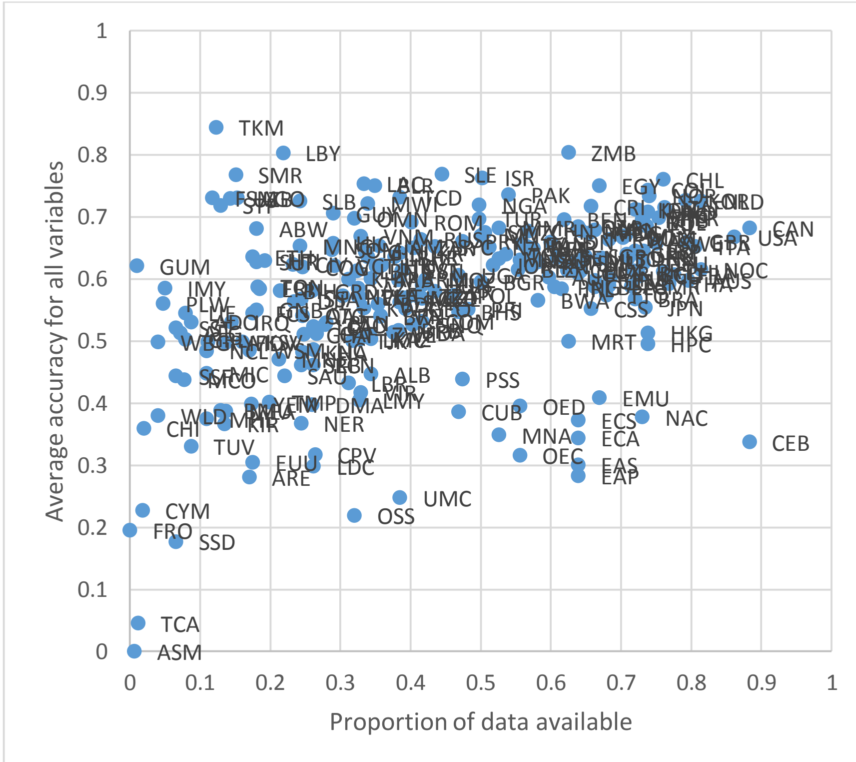


Figure 16: Scatter plot showing the correlation between the amount of data available, and the test accuracy of the networks trained with such data, with the country code highlighted for each point. Note that, ignoring outliers like American Samoa (ASM), the points form an almost horizontal spread between 0.5 and 0.8.

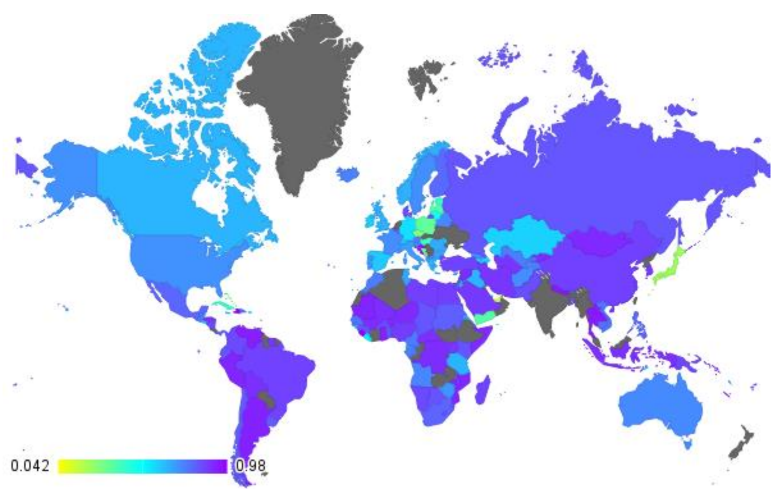


Figure 17: Geographic distribution of accuracy results for Gross Capital formation

Excepting the Middle East and North Africa, the accuracy of "Wage and salaried workers" is average for most countries. Again, this makes sense when considering that almost all values for this variable are missing for this region (see Appendix 1.1 starting on page 64).

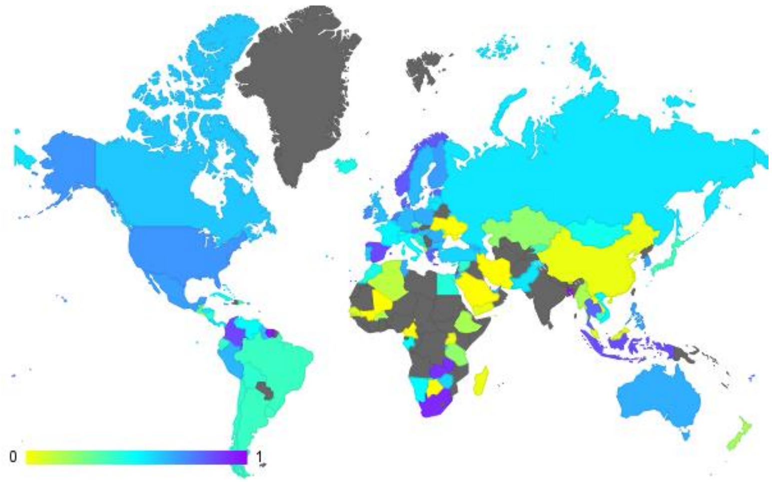


Figure 18: Geographic distribution of accuracy results for Wage and salaried workers

Interestingly, the variable with the absolute lowest amount of data available (Compensation of Employees) was predicted with an average accuracy of 0.6181. This variable is known to be easily connectable—Smets and Wouters's original model gives it the highest degree after GDP, and both the reproducibility experiment and the networks in Appendix 4.1 (starting on page 87) discovered multiple connections for it—indicating that the methodology is able to provide very high accuracies provided a sufficient number of connectable variables exist in the domain knowledge model even with a large proportion of the data missing. The geographic distribution for this and the other variables can be seen in Appendix 5.1 starting on page 112.

6.1.3. Matching network experiment

Three networks were found to be shared among several countries.

6.1.3.1. Network 1

The first network was found to be shared by Fiji and St. Lucia. Both are small island nations and former British colonies with diversified economies (Hausmann, et al. 2011). Because the economies are small, it makes sense for the influence of capital investment to be much more direct on worker compensation and that this capital be what drives the economy (Figure 19)

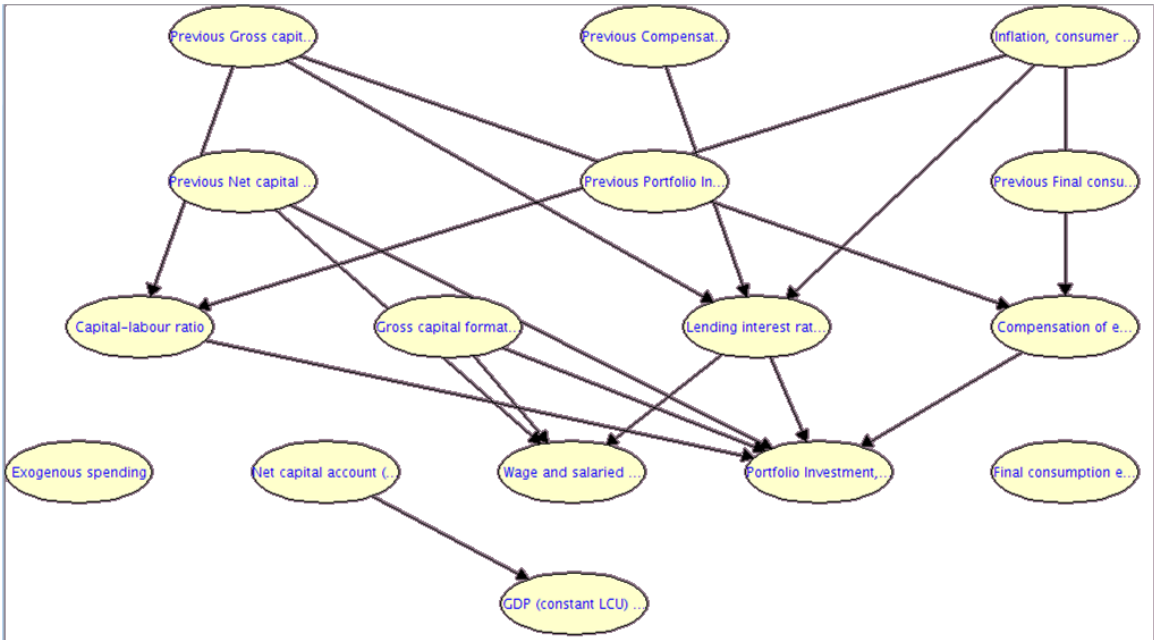


Figure 19: Bayesian network generated for Fiji and St. Lucia

The official World Bank country grouping puts the countries together in 6 of the 11 groups they belong to. The network seems to be a slightly better fit for Fiji, the bigger of the two countries, but this may just be because it has more data available (Table 11).

Table 11: Comparison of groups and accuracies for the countries that share Network 1

Groups shared by the countries	*Country*	*Groups not shared by the countries*	*Data available*	*Average accuracy*
• IDA & IBRD total • Low & middle income • Middle income • Small states • Upper middle income • World	Fiji	• East Asia & Pacific (all income levels) • East Asia & Pacific (developing only) • IBRD only • Pacific island small states	0.3393	0.6359
	St. Lucia	• Caribbean small states • IDA blend • IDA total • Latin America & Caribbean (all income levels) • Latin America & Caribbean (developing only)	0.2619	0.5232

6.1.3.2. Network 2

The second network was found to be shared by Eritrea and Somalia. At first glance, the match seems coincidental due to the lack of data (Figure 20); however, Eritrea and Somalia are geographical neighbors which, according to the World Bank, share all but one of their 13 groups (Table 12). Perhaps most strikingly is the suggestion that exogenous spending drives the economy; however, given that they are also "Fragile and conflict affected situations", it does not seem so surprising.

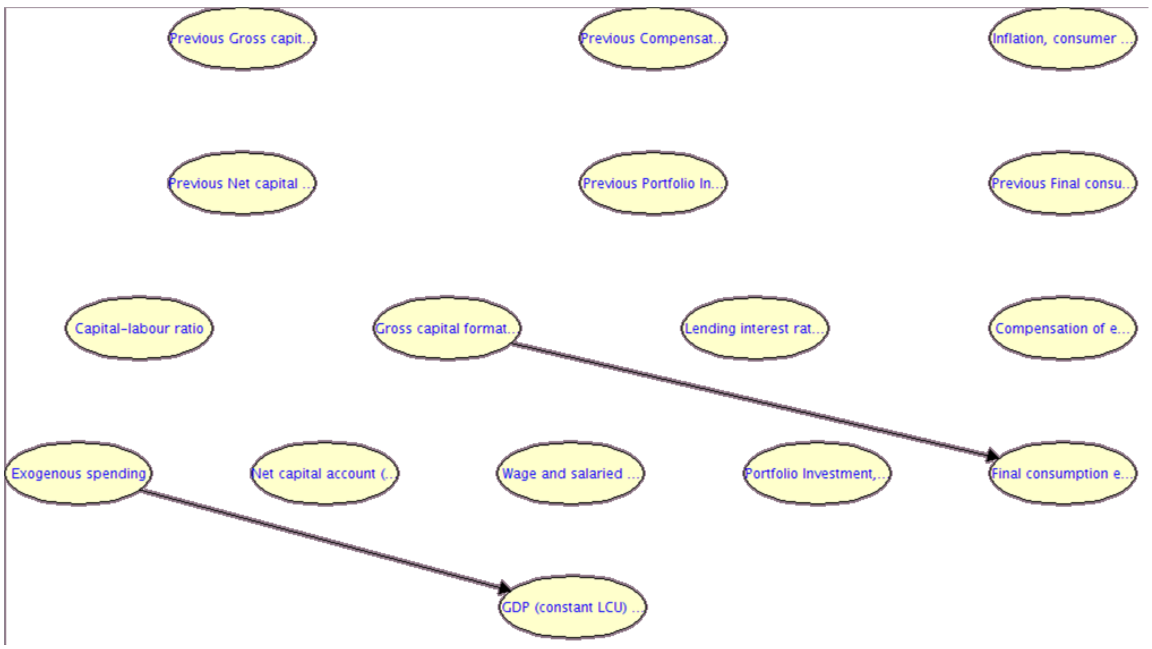


Figure 20: Bayesian network generated for Eritrea and Somalia

Table 12: Comparison of groups and accuracies for the countries that share Network 2

Groups shared by the countries	*Country*	*Groups not shared by the countries*	*Data available*	*Average accuracy*
• Africa • Fragile and conflict affected situations • Heavily indebted poor countries (HIPC) • IDA & IBRD total • IDA only • IDA total • Least developed countries • Low & middle income • Low income • Sub-Saharan Africa	Eritrea		0.1845	0.5842
	Somalia	• Arab World	0.2876	0.6479

6.1.3.3. Network 3

The third network was found to be shared by Iraq, Guyana, and Tonga. Like in Network 1, capital expenditure and portfolio investment seems to play an important role in the economy (Figure 21), however not many other relations could be found. All three nations are known to be developing (Table 13) and to have repeated problems with other countries (Iran in the case of Iraq, Venezuela in the case of Guyana, and the United Kingdom in the case of Tonga) so it is possible that their economies truly are in this state. This is supported by the fact that the model seems to be an above-average fit for all three countries.

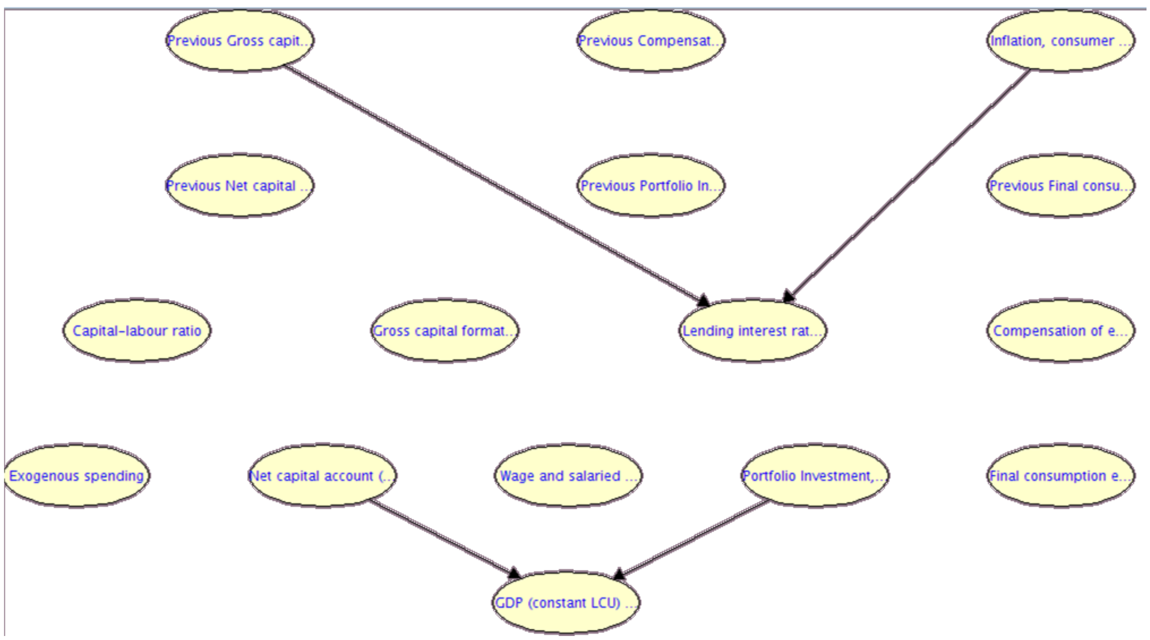


Figure 21: Bayesian network generated for Iraq, Guyana and Tonga

6.1.4. Benchmarking experiment

6.1.4.1. Construction time

Most Bayesian networks are constructed between two and three milliseconds, with the distribution of times extending to 21 milliseconds (Figure 22). This variation in time is not dependent on the amount of data available (correlation coefficient of 0.3589); rather, it shows a quadratic correlation with the number of arcs generated (Figure 23), which is consistent with Friedman et al. This correlation is so much better that the correlation coefficient, even though not designed to detect quadratic correlation, still reports a correlation of 0.7475.

Table 13: Comparison of groups and accuracies for the countries that share Network 2

Groups shared by the countries	*Country*	*Groups not shared by the countries*	*Data available*	*Average accuracy*
• IDA & IBRD total • Low & middle income • World	Iraq	• Arab World • Fragile and conflict affected situations • IBRD only • Middle East & North Africa (all income levels) • Middle East & North Africa (developing only) • Middle income • Upper middle income	0.1448	0.5310
	Guyana	• Caribbean small states • Heavily indebted poor countries (HIPC) • IDA & IBRD total • IDA only • IDA total • Latin America & Caribbean (all income levels) • Latin America & Caribbean (developing only) • Low & middle income • Lower middle income • Middle income • Small states	0.2897	0.7058
	Tonga	• East Asia & Pacific (all income levels) • East Asia & Pacific (developing only) • IDA only • IDA total • Middle income • Pacific island small states • Small states • Upper middle income	0.1825	0.5880

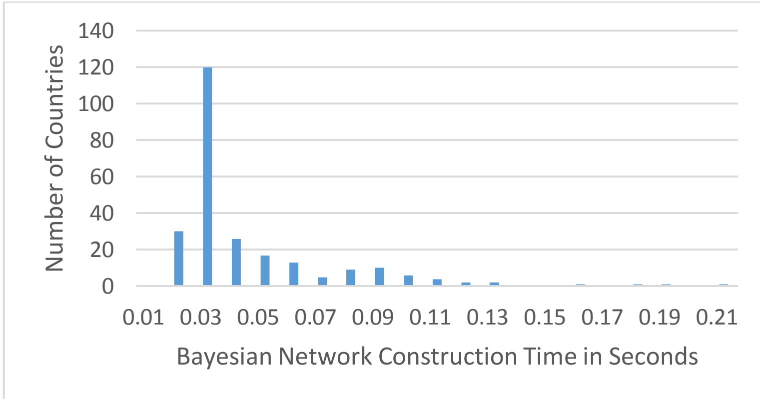


Figure 22: Histogram for the average construction time to build a network using the Smets and Wouters Domain Knowledge model (average over the 20 random data splits)

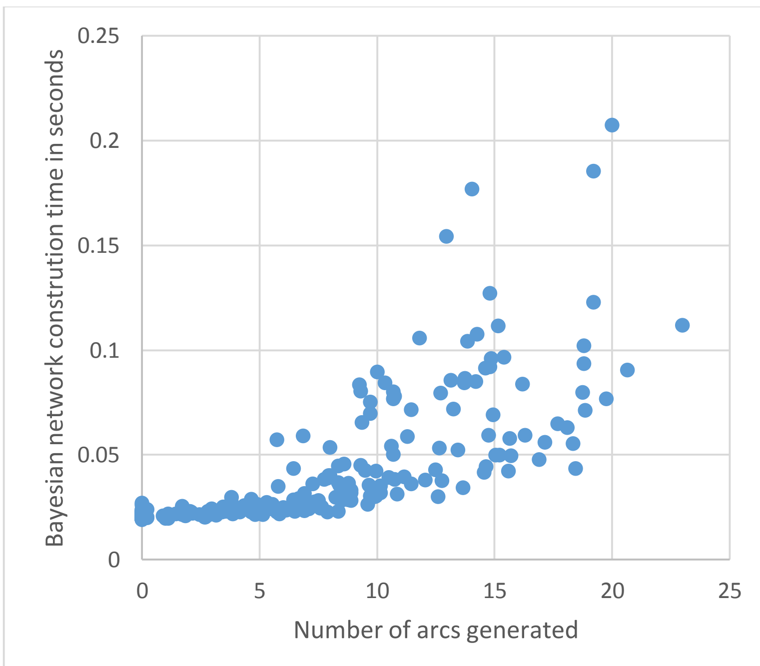


Figure 23: Scatter plot showing the correlation between the number of arcs generated, and the time it took to generate them, using the Smets and Wouters Domain Knowledge model

6.1.4.2. Training time

Training time takes almost no time at all, although there is a slight variability (Figure 24). The time spent in training is not strongly correlated to the amount of data (correlation index of 0.1874). It is better correlated to the number of arcs (correlation index of 0.4902) and might be a function of it (dependency of 0.7580). Again, this is consistent with Friedman et al. The nature of the correlation is shown in Figure 25.

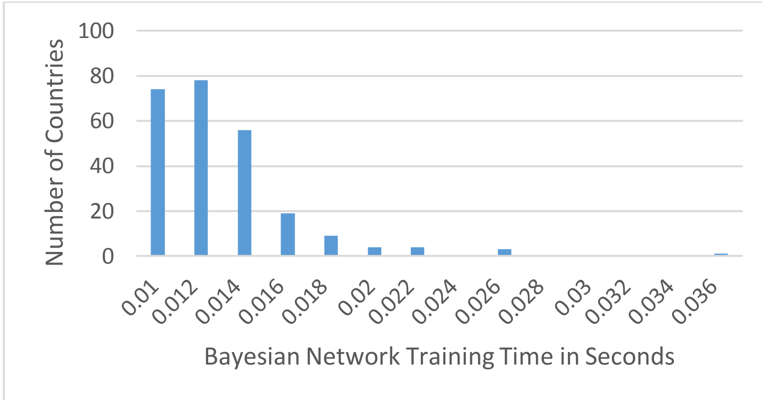


Figure 24: Histogram for the average time to train a network built using the Smets and Wouters Domain Knowledge model (average over the 20 random data splits)

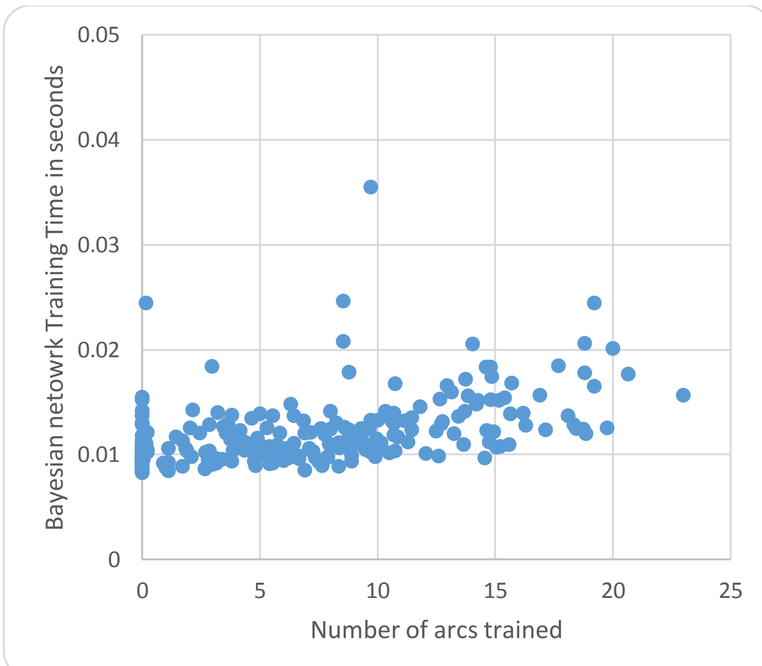


Figure 25: Scatter plot showing the correlation between the number of arcs in the generated network, and the time it took to train them, using the Smets and Wouters Domain Knowledge model

6.1.4.3. Processing time

The total processing time, including all overheads, is distributed in a bell shape centered around 0.4623 seconds, with an amazing 40 countries reporting under 15 milliseconds (Figure 26). It has a 0.5409 correlation coefficient with the amount of data; however, the correlation coefficient with the number of arcs is a much higher 0.8591. It might, in fact, be a function of the number of arcs, with a dependency of 0.7427. This correlation can be seen in Figure 26.

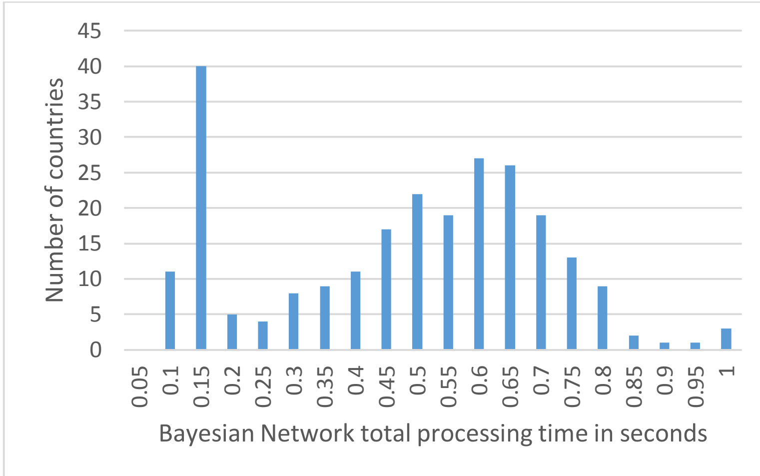


Figure 26: Histogram for the average time to fully process a data split, including network construction and training, using the Smets and Wouters Domain Knowledge model

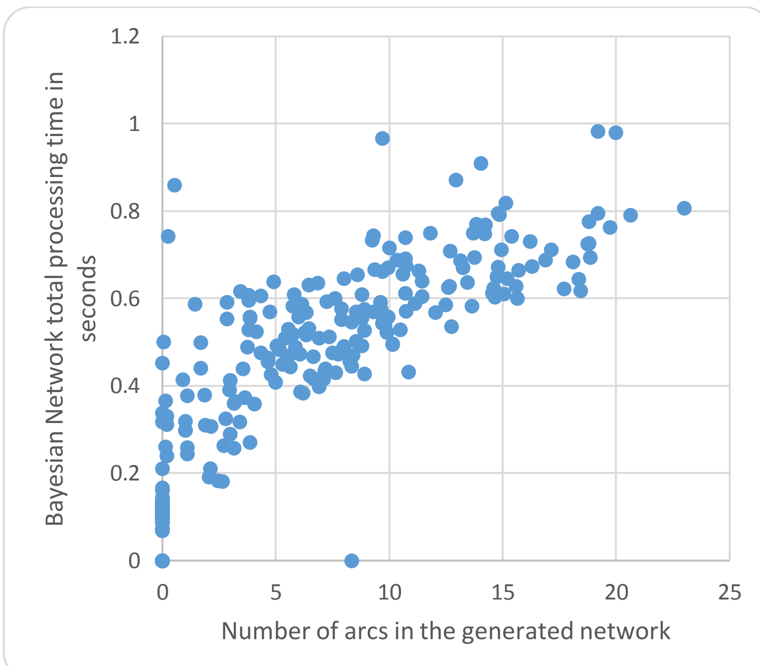


Figure 27: Scatter plot showing the correlation between the average number of arcs in the generated networks (average over the data splits), using the Smets and Wouters Domain Knowledge model

Closer inspection of the results revealed that 15 milliseconds was the time required to discard all arcs: the 40 countries and regions in this range produce networks with no arcs (all variables uniformly randomly distributed). These are mainly those with too little data (such as American Samoa) or which are aggregates of too many diverse countries for the relations to be meaningful (such as the "Low & middle income" economic group).

6.2. Networks generated using UNESCO domain knowledge

The number of arcs generated seems to vary widely for each country. As with the networks built using the Smets and Wouters Domain Knowledge model, it seems influenced by the amount of data available (correlation 0.5314). But, as before, it does not define a function (dependency 0.2567)—at best it could estimate the maximum number of arcs (Figure 28).

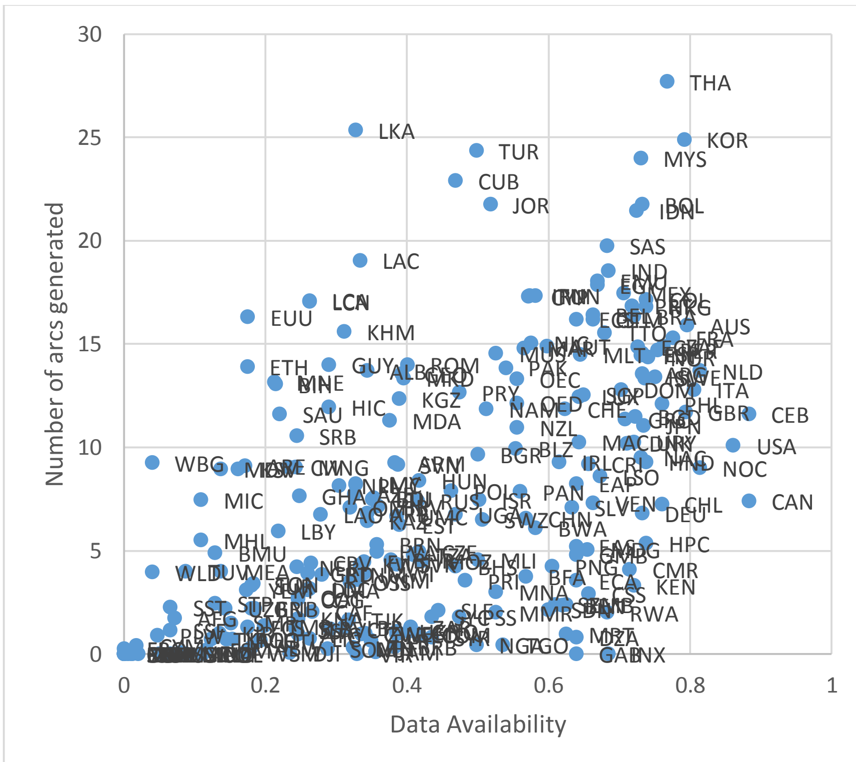


Figure 28: Scatter plot showing the correlation between the amount of data available, and the number of arcs generated, on average, with such data when using the UNESCO Domain Knowledge model. The country code highlighted for each point.

6.2.1. Reconstruction experiment

The average accuracy for all variables in all countries was 0.5405, which represents a 62.15 percent over choosing a discretized bin at random (accuracy of one-third). For most variables, more countries had accuracies above the mean than below it; i.e.: the population of accuracies was denser in the higher quartiles—this can be observed in Table 14. The distribution of the accuracies is left-tailed, which can be evidenced by the interquartile range. This distribution is maintained for most geographic regions (Figure 29) and economic groups (Figure 30). "Latin America & Caribbean" (40 members) and the "High income: nonOECD" (47 members) have slightly different distributions, with their bell curves being centered at slightly lowered values. These two groups share 20 members in common, among which are atypical economies, such as Argentina and Venezuela with their hyperinflation histories, and the Cayman Islands with their tax haven economy. These situations make them harder to predict (the average accuracies for all variables of these three countries are, respectively, 0.2361, 0.1846, and 0.0824) and are probably what skews the distribution for this region and this group.

Table 14: Interquartile ranges for the distribution among all countries of the average test accuracies (average among the 20 random splits)

Variable name	*Avg*	*Min*	*Q1*	*Med*	*Q3*	*Max*
Labor force with primary education	0.5093	0.0000	0.3561	0.5333	0.7406	0.9608
Labor force with secondary education	0.5299	0.0000	0.3753	0.5763	0.7502	0.9817
Labor force with tertiary education	0.4736	0.0000	0.2447	0.5417	0.6750	0.9762
Scientific and technical journal articles	0.5786	0.0000	0.4554	0.6003	0.7082	0.9458
Trademark applications, total	0.6340	0.0238	0.5410	0.6637	0.7727	0.9789
General government final consumption expenditure	0.5518	0.0789	0.4504	0.5580	0.6500	0.9286
Net official development assistance and official aid received	0.5857	0.0000	0.4717	0.5786	0.6987	0.9866
Agriculture, value added	0.5422	0.0439	0.4389	0.5460	0.6538	0.9533
Industry, value added	0.5218	0.0000	0.4291	0.5461	0.6432	0.9616
Manufacturing, value added	0.5170	0.0000	0.3967	0.5218	0.6665	0.8884
Services, etc., value added	0.5600	0.0500	0.4218	0.5810	0.6986	0.9129
Unemployment, total	0.4975	0.1367	0.3689	0.5000	0.6167	0.9608
GDP growth	0.5680	0.0702	0.4636	0.5550	0.6696	0.9717
GDP per capita, PPP	0.5279	0.0417	0.4383	0.5492	0.6344	0.8950
Previous GDP growth	0.5822	0.1146	0.4670	0.5759	0.7232	0.9491
Previous GDP per capita, PPP	0.5073	0.0588	0.3806	0.5258	0.6325	0.9075
Average for all variables	*0.5405*	*0.1210*	*0.4919*	*0.5423*	*0.5980*	*0.9286*

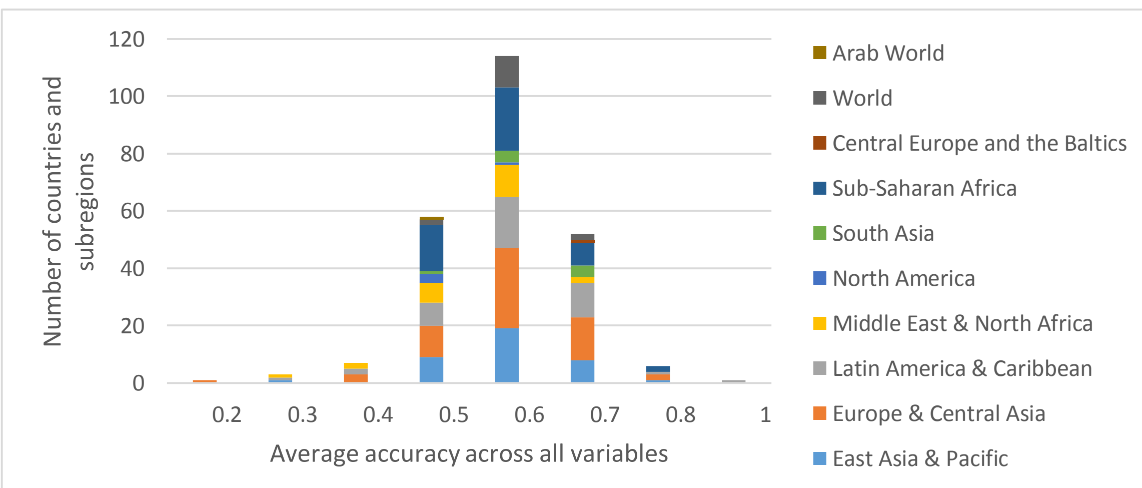


Figure 29: Histogram of average accuracies for all variables for the UNESCO Domain Knowledge Model Reconstruction Experiment, with each country's main geographic region highlighted.

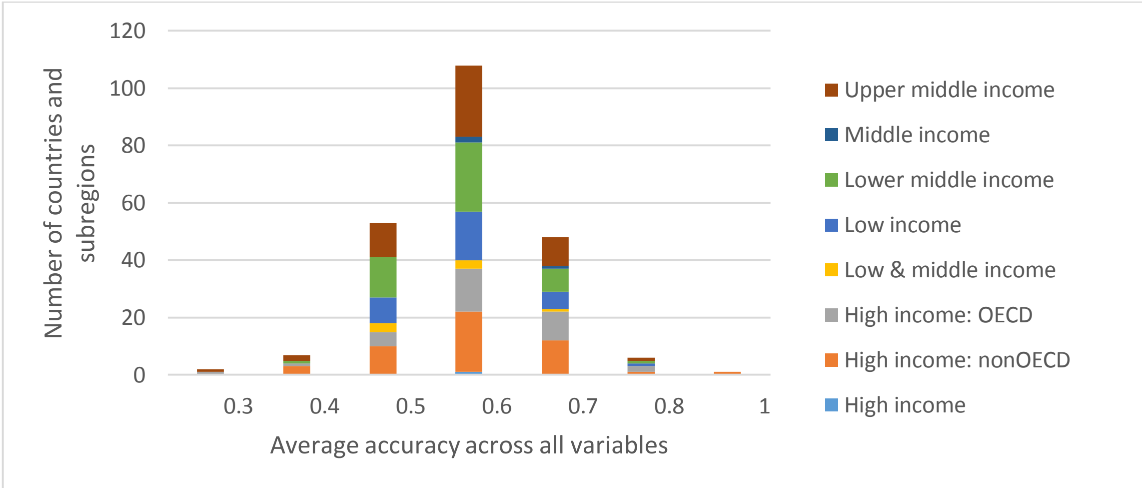


Figure 30: Histogram of average accuracies for all variables for the UNESCO Domain Knowledge Model Reconstruction Experiment, with each country's main economic group highlighted.

The bell shape is typical for generalizable models. With a correlation coefficient of 0.1750, the different accuracies cannot be attributed to the available data (see Figure 31). Rather, they seem to be a true indication to how well the model adapts to the countries. Again, this suggests good tolerance to missing data.

The variable with the highest average accuracy was "Trademark applications" with 0.6340, representing an improvement of 90 percent over choosing the discretized bin at random. There are 107 countries above this average (shown in dark blue and violet in Figure 32) and only 16 countries for which the average accuracy is worse than random (shown in yellow and green in Figure 32). The highest accuracies (violet in in Figure 32) are found in south and east Europe and the East Indies, suggesting that, for these countries. These areas are not exceptional for their data availability (see Appendix 1.2 starting on page 68), and indeed, the global availability of trademark data is average (see Figure 7 in Chapter 4; page 14) suggesting that, for those countries, trademarks truly are much more strongly related to the production and economic variables, and that the algorithm was indeed able to discover these relations.

The variable with the lowest average accuracy was "Labor force with tertiary education" with 0.4736, representing a 41 percent improvement over choosing the discretized bin at random. This makes sense when considering that it has the lowest amount of data available (see Figure 7 in Chapter 4; page 14) and indeed, all the countries with worse-than-random accuracy (shown in yellow and green in Figure 32) have less than a quarter of the data available (see Appendix 1.2 starting on page 68).

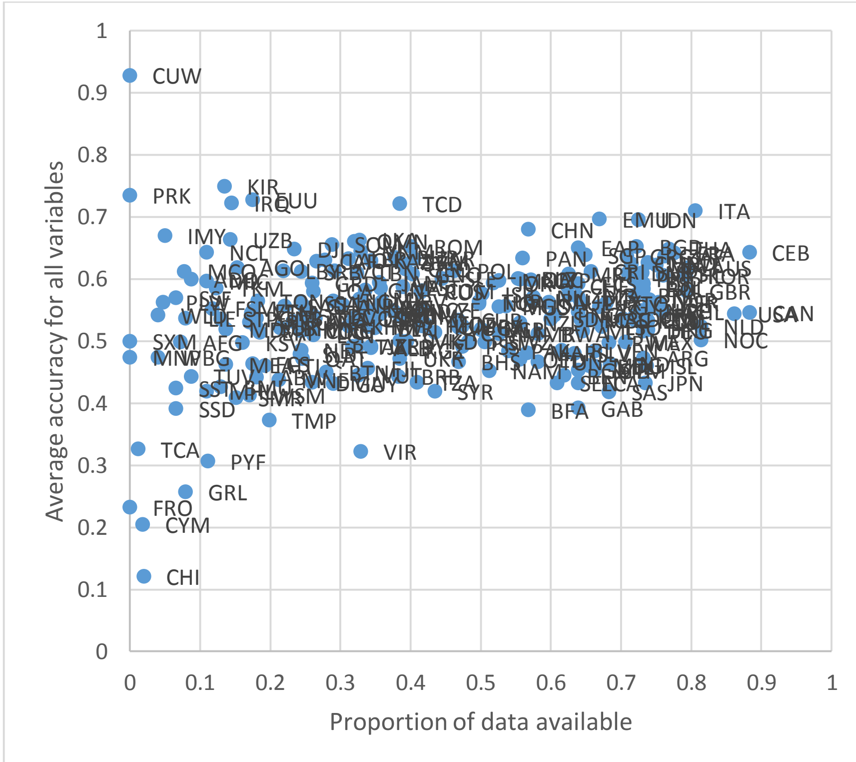


Figure 31: Scatter plot showing the correlation between the amount of data available, and the test accuracy of the networks trained with such data, with the country code highlighted for each point. Note that, ignoring outliers like The Channel Islands (CHI), the points form an almost horizontal spread between 0.4 and 0.7.

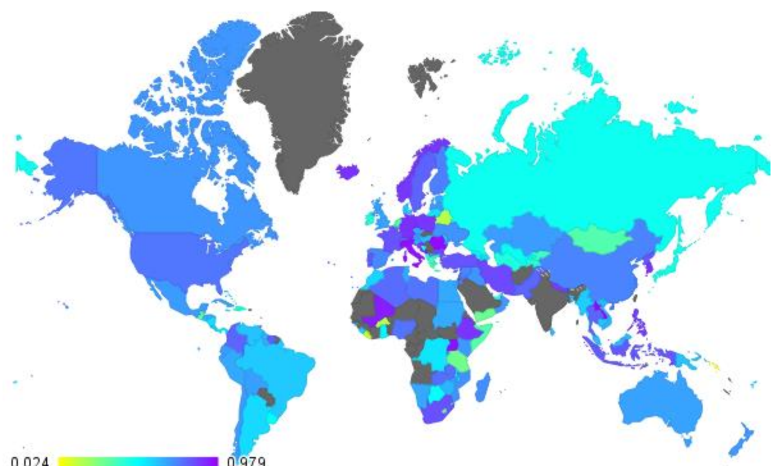


Figure 32: Geographic distribution of accuracy results for Trademark Applications

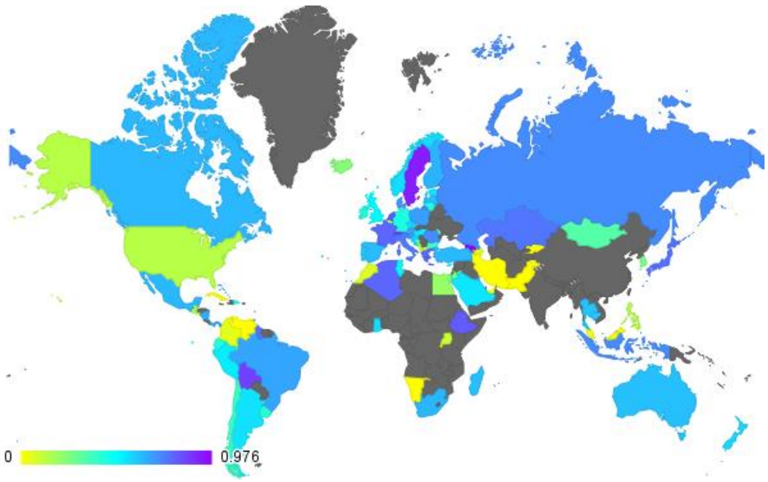


Figure 33: Geographic distribution of accuracy results for Labor force with tertiary education

It is of note that the other two variables with this amount of data available ("Labor force with primary education" and "Labor force with secondary education") obtained much higher accuracies, with their first quartile being better than random (see Table 14). This suggests that these variables have a much stronger relation with the categories they connect to, even in the countries where tertiary education could not be estimated. The geographic distribution for these and the other variables can be seen in Appendix 5.2 starting on page 118.

6.2.2. Matching network experiment

Two networks were found to be shared.

6.2.2.1. Network 1

The first network was found to be shared by the Republic of the Congo and Gabon. These are two geographical neighbors that share 8 out of 12 groups (Table 15). Both countries have no data for the education variables or for the number of trademarks (see Appendix 2.2 starting on page 79), however, for all others, relations were found (Figure 34). As is known for these countries, official development aid is an important factor, and this is reflected by the network.

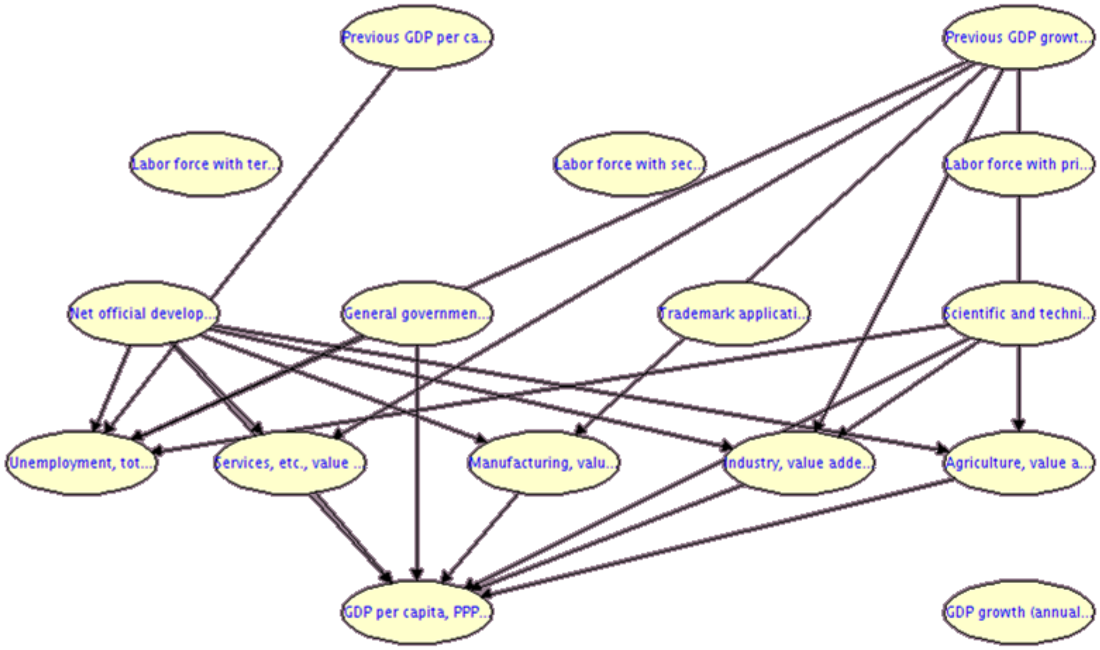


Figure 34: Bayesian network generated for the Republic of the Congo and Gabon

Table 15: Comparison of groups and accuracies for the countries that share Network 1

<table>
<tr><th>Groups shared by the countries</th><th>Country</th><th>Groups not shared by the countries</th><th>Data available</th><th>Average accuracy</th></tr>
<tr><td rowspan="2">• Africa
• IDA & IBRD total
• Low & middle income
• Middle income
• Sub-Saharan Africa (all income levels)
• Sub-Saharan Africa (developing only)
• Upper middle income
• World</td><td>Congo</td><td>• IDA blend
• IDA Total
• Lower & middle income
• Heavily indebted poor countries (HIPC)</td><td>0.2461</td><td>0.5177</td></tr>
<tr><td>Gabon</td><td>• IBRD only
• Middle income
• Small states
• Other small states</td><td>0.6389</td><td>0.3924</td></tr>
</table>

6.2.2.2. Network 2

The second network was found to be shared between two regional aggregates: “Sub-Saharan Africa (all income levels)” and “Sub-Saharan Africa (developing only)”. Obviously, the latter is a sub-region of the former, and suggests that the Sub-Saharan developed countries, when aggregated with their neighbors, do not behave differently from the developing countries. Checking the definition (World Bank n.d.), reveals that there are three developed countries in sub-Saharan Africa, and 46 developing countries. This match, therefore, stems simply from the fact that the developing countries far outweigh their developed neighbors when aggregated, and indeed is a slightly better match for just the developing countries than when the developed countries are included (Table 16).

The network is extremely similar to Network 1, which makes sense considering that the countries that Network 1 was made for are members of both of these regions. The main difference is the much lower number of dependency relations found for the GDP per capita, which is probably due to the variations caused by other members.

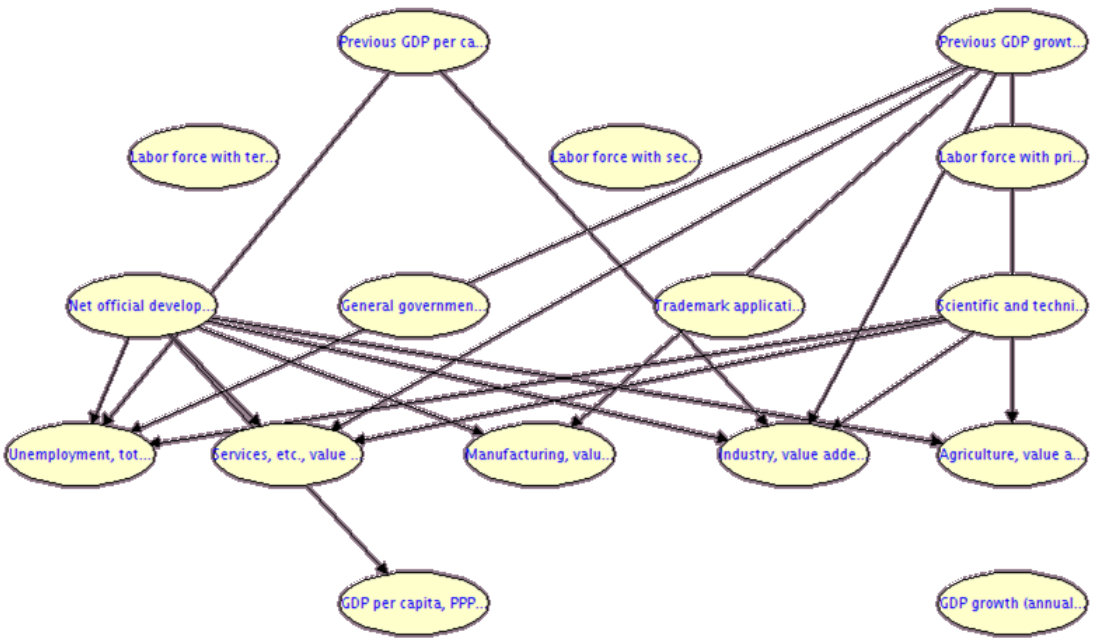


Figure 35: Bayesian network generated for Sub-Saharan Africa (all income levels) and Sub-Saharan Africa (developing only)

Table 16: Comparison of groups and accuracies for the countries that share Network 2

Superregions shared by regions	*Regions*	*Superregions not shared*	*Data available*	*Average accuracy*
• Africa • World	Sub-Saharan Africa (all income levels)		0.5472	0.5616
	Sub-Saharan Africa (developing only)	Low & middle income	0.5472	0.5696

6.2.3. Benchmarking experiment

6.2.3.1. Construction time

Most Bayesian networks are constructed between two and four milliseconds, with the distribution of times extending to 19 milliseconds (Figure 36). This variation in time is not dependent on the amount of data available (correlation coefficient of 0.2804); rather, it shows a slight quadratic correlation with the number of arcs generated (Figure 37), which is consistent with Friedman et al. This correlation is so much better that the correlation coefficient, even though not designed to detect quadratic correlation, still reports a correlation of 0.6847.

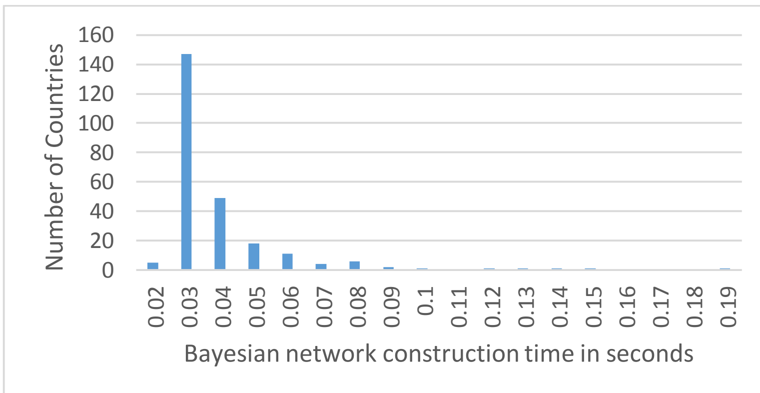


Figure 36: Histogram for the average construction time to build a network using the UNESCO Domain Knowledge model (average over the 20 random data splits)

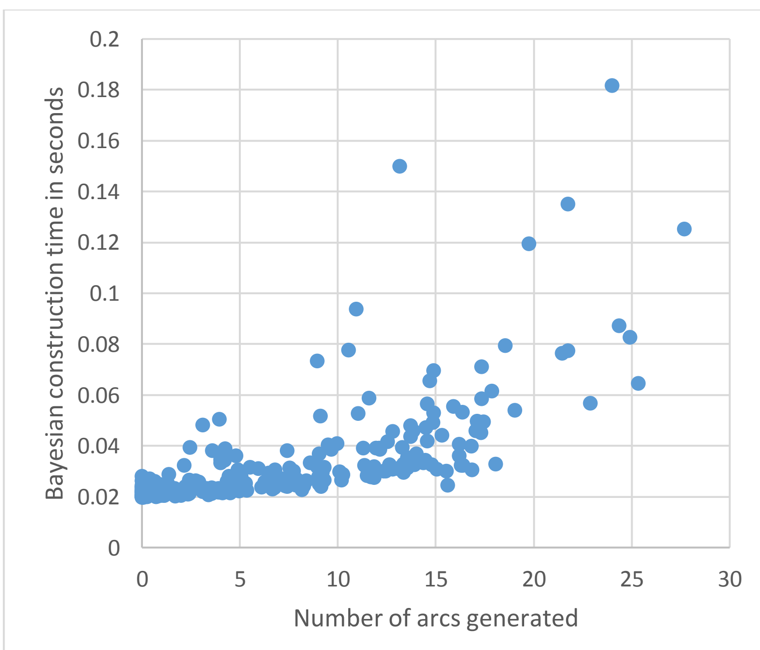


Figure 37: Scatter plot showing the correlation between the number of arcs generated, and the time it took to generate them, using the UNESCO Domain Knowledge model

6.2.3.2. Training time

Training time takes almost no time at all, although there is a slight variability (Figure 38). The time spent in training is not strongly correlated to the amount of data (correlation index of 0.0079). It is better correlated to the number of arcs (correlation index of 0.2261) and might be a function of it (dependency of 0.7205). Again, this is consistent with Friedman et al. The nature of the correlation is shown in Figure 39.

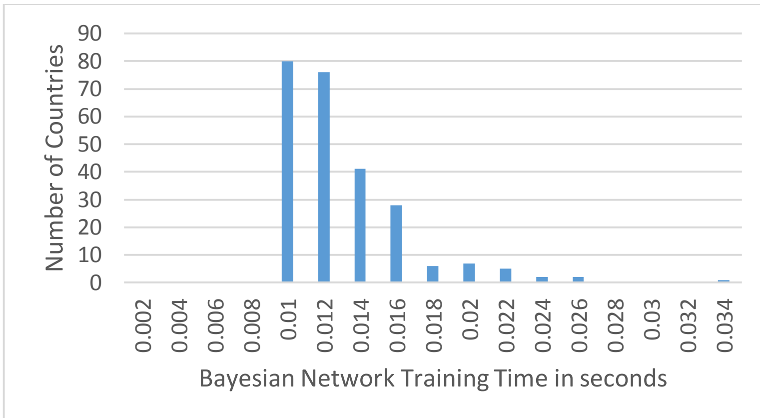


Figure 38: Histogram for the average time to train a network built using the UNESCO Domain Knowledge model (average over the 20 random data splits)

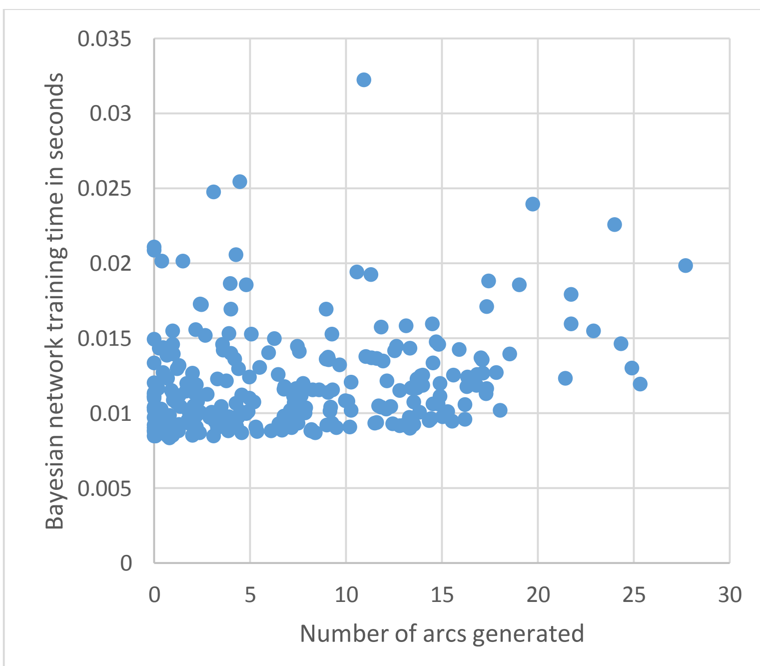


Figure 39: Scatter plot showing the correlation between the number of arcs in the generated network, and the time it took to train them, using the UNESCO Domain Knowledge model

6.2.3.3. Processing time

The total processing time, including all overheads, is distributed in a bell shape centered around 0.5171 seconds (Figure 40). It has a 0.5502 correlation coefficient with the amount of data; however, the correlation coefficient with the number of arcs is a much higher 0.7409. It might, in fact, be a function of the number of arcs, with a dependency of 0.7995. This correlation can be seen in Figure 41.

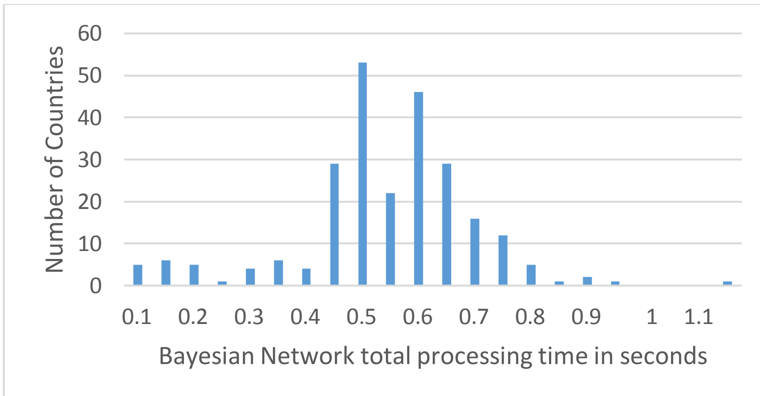


Figure 40: Histogram for the average time to fully process a data split, including network construction and training, using the UNESCO Domain Knowledge model

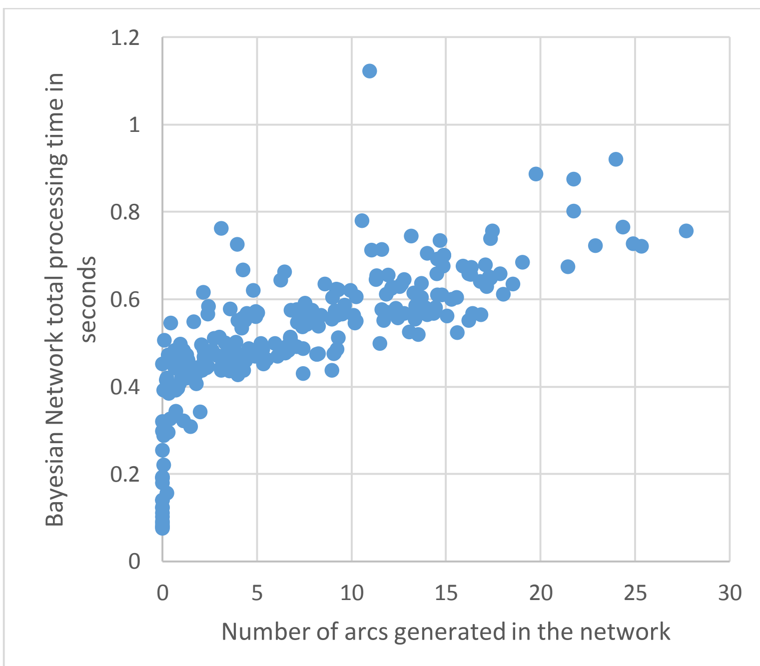


Figure 41: Scatter plot showing the correlation between the average number of arcs in the generated networks (average over the data splits), using the Smets and Wouters Domain Knowledge model

6.3. Comparison

For both models, the number of arcs appears to be uniformly distributed, with a function of the data available as its upper bound. The number of arcs both models create for the same country has a slight correlation (correlation index of 0.3306) which cannot be explained alone by the amount of data (the correlation index of data availability for the variables of both models is of 0.0537). This suggests that the models do indeed capture latent relations about how the countries function. However, it does not mean that the number of arcs a country gets in one model can be used to estimate the number of arcs in the other: the dependency score for Smets and Wouters arcs → UNESCO arcs is only 0.1431, and the dependency score for UNESCO arcs → Smets and Wouters arcs is only 0.2159. At best, the distribution among several countries for one model can be used to estimate a bound (Figure 42)

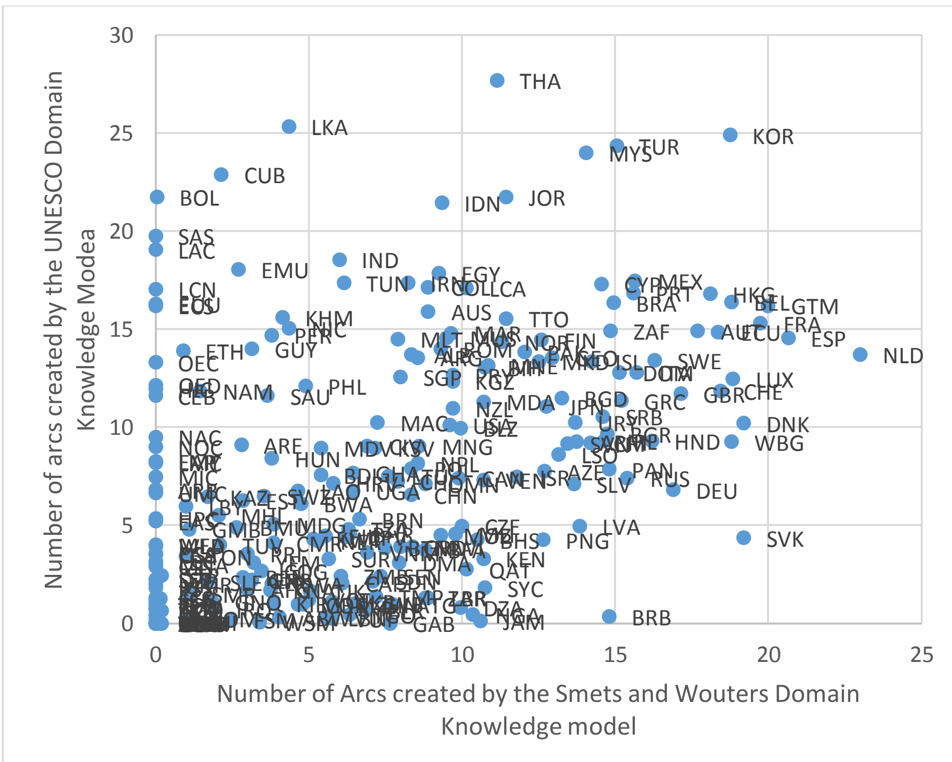


Figure 42: Scatterplot showing the correlation between the number of arcs generated on average, using the Smets and Wouters domain knowledge model and the UNESCO domain knowledge model. The country code is highlighted for each point.

6.3.1. Reconstruction experiment

The average accuracy for all variables in all countries was similar for both models (three percentage points of difference) and resulted in more countries with accuracies above this average than below it. The distribution is generally maintained within geographic regions and economic groups.

Although the distribution for the UNESCO model is centered at a slightly lower value, the distribution is narrower suggesting better generalizability, reliability, and stability. The UNESCO model is also less susceptible to missing data: its correlation coefficient (accuracy vs data available) is less than half what it is for the Smets and Wouters Domain Knowledge model. This makes the UNESCO model more attuned to the initial goal of improving tolerance to missing values.

6.3.2. Benchmarking experiment

The network construction time does not differ significantly between models. This makes sense since both models have similar numbers of variables, categories and arcs between them. The UNESCO Domain Knowledge model, which has slightly more arcs between its categories, creates slightly more arcs on average, which, as expected, takes quadratically more time. Both construction times appear to be quadratically correlated to the number of arcs generated, as in Friedman et al.

Training time for the networks generated by both models is very short (less than 4 milliseconds) and seems to be a function of the number of arcs generated, rather than of the amount of data (both dependencies are above 0.7).

The total processing time for both networks was randomly distributed centered around half a second, and was much more strongly correlated to the number of arcs than to the amount of data available for both models.

7. CONCLUSIONS

Interpretability serves an important role in going from the age of information to the age of knowledge (Tan, Steinbach and Kumar 2005). It is very easy to make bold claims about the world, even when these are well informed, when there is no way to test them. The method presented herein provides a way to test them, creating new knowledge. However, new knowledge creates new information, which then needs to be further processed to lead to understandable knowledge—the original goal of interpretability (Liu, Cocea and Gegov 2016). This is the challenge faced in this project: when trying to introduce interpretability, knowledge is generated, and with it, its associated data.

We were able to create logical models for all countries and regions using our methodology. The resulting models are consistent with existing knowledge known about the regions during the years covered by the data. A domain knowledge representation was created based on Smets and Wouters original model. However, the domain knowledge model created describes many more relations than the original Smets and Wouters model. These relations are more easily detectable than the ones originally proposed by Smets and Wouters—possibly indicating that the expert intuition is not supported as strongly by statistics, which would make the new relations stronger and better. It is possible that, with different minimum and threshold parameters the original arcs can be captured, but a more granular domain knowledge model may be needed if the other relations discovered are truly not desired.

The Smets and Wouters Domain Knowledge model proved an adequate fit (average accuracy above 0.5 for all variables) for 75 percent of the countries: the average First Quantile accuracy was 0.5113, suggesting that the networks generated for most countries were able to estimate most of the variables adequately. The network generated for the case study of the United States also proved to be better in the variables Smets and Wouters chose not to report, again suggesting that the generated model does a better job of giving all variables reasonable accuracies

The UNESCO Domain Knowledge model proved an adequate fit (average accuracy above 0.5 for all variables) for 70 percent of the countries. This is beyond reasonable, given that the UNESCO model originates from abstract concepts and handpicked variables which attempt to represent those concepts. Even in countries outside of that 70 percent, the networks generated were still able to estimate most of the variables adequately, as evidenced by the average First Quantile accuracy of 0.4919. This indicates that some of the variables chosen simply are harder to estimate (mostly because of the large amount of missing data) which, given the proposed methodology, can easily be compensated by adding alternate measures of that concept into the same category.

The UNESCO model, because of its more comprehensive nature, also generalizes better than Smets and Wouters' Domain Knowledge model to more countries. It has a higher average minimum accuracy for all variables at 0.1210 and a higher average maximum accuracy at 0.9286.

When the same network was generated for several countries or regions, the countries or regions, they were, in fact, found to be related. This suggests that countries can be clustered by their network and yield meaningful results. It also supports the claim that the network is capturing latent relations in the data, reflecting how the countries truly function.

The complexity of the networks also remained reasonable: although 52 arcs were possible under the Smets and Wouters Domain Knowledge model, and 57 arcs were possible under the UNESCO Domain Knowledge model, both models produced on average seven arcs per network, and never more than 28. This means the networks are easy to read and interpret (Liu, Cocea and Gegov 2016, García, et al. 2009) without producing results worse than random, as was originally set out.

The algorithm is also tolerant to missing data, as was originally set out: all correlation coefficients comparing accuracies to the amount of data available are less than 0.5, with the UNESCO coefficients being lower, and thus, better. This is because the model is good at discovering semantically sensible connections even with little data, and given a sufficient number of connectable variables, each variable will always yield a good accuracy. This further supports the claim that the algorithm can discover true relationships in the data.

The algorithm is efficient in comparison to existing methods. The time was found to be more strongly related to the complexity of the Domain Knowledge model than to the amount of data. The processing time being centered around half a second suggests that, for domain knowledge models of this complexity, most of the processing time is overhead. This overhead is expected to remain relatively constant as the complexity grows.

Ultimately, the project succeeds in providing a tool for the creation of useful knowledge, the quantification of existing knowledge, and the promotion of the pursuit of knowledge.

7.1. Future Work

The method currently assumes that assigning the variables to their categories is a manual process, even though, in the data used, the variables do have metadata associated with them. This metadata indicates, among other things, a category for the variable, and is constantly being revised (World Bank 2015). Such practice is increasingly common (Ramezani, et al. 2010) so a technique to take advantage of it seems useful. It is therefore posited that downloading all 1,421 of the variables of the World Bank, automatically detecting their categories, and using a suitable Domain Knowledge model with this method, could yield a comprehensive economic model; and that by adjusting the thresholds of the system, the truly most important variables could be discovered.

In the Reconstruction experiment, data from years following the year being tested could have been used in the training, mainly to reduce the impact of missing data (when this missing data occur in several successive years). Experiments to find a true prediction accuracy would prove extremely valuable, and could give a better indication on the use of this method to develop policy-setting models, and thus truly replace Smets and Wouters (The European Central Bank n.d.).

The work done herein is also done in only area: the area of economics. An interesting study could arise by trying data from several areas—for example, a genomics application could have several gene groups as the categories and help narrow down which individual genes truly relate to one another and in what ways. Because the method is not confined to Bayesian networks, other implementations are worthwhile exploring—for example, an implementation for a rule-based decision tree could be used in psychology to discover the thresholds it takes for environmental and biological variables to trigger an emotional state and which variables these may be; an implementation for deep learning could aid in computer vision to decisively interpret what each layer of the neural net is actually learning, and thus aid in designing better, more powerful networks more quickly.

Accuracy computations currently do not account for situations when more than one outcome has the same, or very similar, probabilities: a real result observed had the “high” bin with probability of 0.414, and of the “medium” bin with probability 0.415. The prediction was marked as incorrect since the expected value was “high”, but “medium” was technically higher. A smarter accuracy measure might be able to detect these cases and mark them as correct or partially correct, yielding to an accuracy measure more reflective of human interpretation of the results.

In the matching network experiment, networks were grouped by perfect equality. Given that all the networks generated have the same nodes, an actual similarity measure is feasible. Once implemented, models could be clustered using traditional clustering methods to identify groups of similarly-functioning economies.

The discretization method was specifically chosen to maximize training efficacy. Other discretization methods, or indeed the use of continuous conditional probabilities in the generated networks, could prove an interesting comparison. A study, of whether changing the point in the data pipeline where discretization is undertaken affects the results, could also be performed. A comparison of performance between discretized and continuous networks could also prove interesting.

An exploration could also be undertaken of the arc evaluation method: although STE is consistent with correlation and mutual information, it requires discarding a data pair if the value is not known for either. The use of another measure, perhaps one that doesn’t have this limitation, might yield better networks.

Finally, the treatment of missing values could be improved without the need to introduce new algorithms: the networks that were found to be good at estimating missing values could very well be used to estimate those missing values, and re-trained with them. The validity of this, however, escapes the scope of this project, and is left as future work.

BIBLIOGRAPHY

Barhoumi, Karim, Szilard Benk, Riccardo Cristadoro, Ard Den Reijer, Audronė Jakaitiene, Piotr Jelonek, António Rua, Gerhard Rünstler, and Karsten Ruth and Christophe Van Nieuwenhuyze. 2008. "Short-term forecasting of GDP using large monthly datasets: a pseudo real-time forecast evaluation exercise." *Occasional Pa p er series* (European Central Bank) (84). https://www.ecb.europa.eu/pub/pdf/scpops/ecbocp84.pdf.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning.* New York: Springer.

Bokova, Irina. 2010. "Foreword." In *Engineering: Issues, Challenges and Opportunities for Development*. Unesco Report.

Bouckaert, Remco. 2015. "Class BayesNet." *Weka doc dev.* Sep 11. http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/BayesNet.html.

—. 2008. "Search Algorithms." *Bayesian Classifiers in Weka for Version 3-5-7.* The University of Waikato. May 12. http://www.cs.waikato.ac.nz/~remco/weka_bn/node8.html.

Cascio, Keith, and the Automated Reasoning Group. 2010. *SamIam.* UCLA. http://reasoning.cs.ucla.edu/samiam/.

Casillas, Jorge, Oscar Cordón, Francisco Herrera, and Luis Magdalena. 2013. "Interpretability Improvementes to find the Balance Interpretability-Accuracy in Fuzzy Modeling: An Overview." In *Interpretability Issues in Fuzzy Modeling*. Springer.

Chickering, D., and Heckerman. 1997. *Microsoft Corporation (assignee). Patent 7,251,636.* http://patentimages.storage.googleapis.com/pdfs/US7251636.pdf.

Cooper, G., and E. Herskovits. 1992. "A Bayesian Method for the Induction of Probabilistic Networks from Data." *Machine Learning* (Springer International.) 9 (4): 309-347.

da Silva, Jo, Susan Thomas, Tony Marjoram, and Ron Watermeyer. 2010. "Engineering, the MDGs and other international development goals." In *Engineering: Issues, Challenges and Opportunities for Development*. UNESCO Report.

Davis Kho, Nancy. 2016. "The State of Big Data." *EContent* (Information Today Inc.).

Del Negro, Marco, and Frank Schorfheide. 2009. "Monetary Policy Analysis with Potentially Misspecified Models." *American Economic Review* 99: pp. 1415-50.

Drahokoupil, Jan. 2013. "Investment incentive." *Encyclopædia Britannica Online*. Encyclopædia Britannica Inc. May 30. http://www.britannica.com/EBchecked/topic/1929166/investment-incentive.

Duda, Richard O, Peter E Hart, and David G Stork. 2001. *Pattern classification.* 2nd ed. New York: Wiley.

Fernandez Macho, F. J., A. C. Harvey, and J. H. Stock. 1987. "Forecasting and Interpolation using Vector Autorregressions with Common Trends." *Annales D'économie et de Statistique* pp. 279-287.

Finnerty, John. 2014. "Education Funding Decision Dogs Gov. Corbett." *The Herald Sharon.* http://www.sharonherald.com/news/education-funding-decision-dogs-gov-corbett/article_42721152-6236-11e4-b5d4-170cc5bcbb52.html.

Friedman, N., I. Nachman, and D., Peér. 1999. "Learning of Bayesian Network Structure from Massive Datasets: The 'Sparse Candidate' Algorithm." *UAI'99Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* 206-215.

Gagliardi Cozman, Fabio. 2001. "JavaBayes: Bayesian Networks in Java." *University of São Paulo.* 1 31. http://www.cs.cmu.edu/~javabayes.

García, S., A. Fernández, J. Luengo, and F. Herrera. 2009. "A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability." *Soft Computing: A Fusion of Foundations, Methodologies and Applications* 13.

Gonzalez, Steven. 2000. "Neural Networks for Macroeconomic Forecasting: A complementary Approach to Linear Regression Models." *Working Papers* (Department of Finance).

Hausmann, R., C. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. S. Jiménez, A. Simoes, and M. Yıldırım. 2011. "The Atlas of Economic Complexity - Mapping paths of prosperity." Edited by Center for International Development. (Harvard University.).

Helman, P., R. Veroff, S Atlas, and C. Willman. 2005. "A Bayesian Network Classification Methodology for Gene Expression Data." *Journal of Computational Biology* 11 (4). Accessed January 20, 2005.

International Labour Organization. n.d. ""Main statistics (annual) - Unemployment"." *LABORSTA Internet.* http://laborsta.ilo.org/applv8/data/c3e.html.

—. n.d. "Indicator 8: Educational Attainment of the Youth Labour Force." *Youth Labour Market Indicators.* Youth Employment Network. http://www.ilo.org/public/english/employment/yen/whatwedo/projects/indicators/8.htm.

ITEP. 2013. "Tax Incentives: Costly for States, Drag on the Nation." ITEP Reports." *Institute on Taxation and Economic Policy.* http://itep.org/itep_reports/2013/08/tax-incentives-costly-for-states-drag-on-the-nation.php.

Jaffe, K., A. Rios, and A. Florez. 2012. "Statistics shows that economic prosperity needs both high scientific productivity and complex technological knowledge, but in different ways." *Interciencia* 38.

Jones, Mervyn. 2010. "Continuing engineering education and professional development." In *Engineering: Issues, Challenges and Opportunities for Development*. UNESCO Report.

Jowitt, Paul. 2010. "Engineering, innovation, social and economic development." In *Engineering: Issues, Challenges and Opportunities for Development*, by UNESCO. UNESCO Report.

Keller, James M., Derong Liu, and David B. Fogel. 2015. *Fundamentals of Computational Intelligence.*

King, Robert G. Mark W. Watson. 2012. "Inflation and Unit Labor Cost." *Journal of Money, Credit and Banking* 111-149.

Liu, Han, Mihaela Cocea, and Alexander Gegov. 2016. "Interpretability of Computational Models for Sentiment Analysis." *Studies in Computational Intelligence* (Springer) 639: pp 199-220.

Lo, Andrew W. 2014. "Macroeconomic Modeling and Financial Stability: Lessons from the Crisis." *The Clearing House*, Oct. https://www.theclearinghouse.org/publications/2014/banking-perspective-q4-2014/macroeconomic-modeling-and-financial-stability.

Microsoft. 2007. *Excel statistical functions: STEYX.* https://support.office.com/article/STEYX-function-4cb00b43-c209-4509-980b-ce4ec8431897.

Migiro, A. 2016. "No 'one size fits all' approach to development." *UN News Center.* June 16. http://www.un.org/apps/news/story.asp?NewsID=35048.

Miles, Ian. 2010. "Engineering, foresight and forecasts of the future." In *Engineering: Issues, Challenges and Opportunities for Development*, by UNESCO. UNESCO.

Millington, Ian, and John Funge. 2009. *Artificial Intelligence for Games.* Second Edition. Morgan Kaufmann.

Mora, J.; Torre, F.; Torre, F. 2013. "Contribución de la enseñanza de la ingeniería a la generación de conocimiento productivo." *Memorias del IV Congreso Iberoamericano de Enseñanza de la Ingeniería* (Asociación Iberoamericana de Instituciones Enseñanza de la Ingeniería.).

Ramezani, Maryam, Hans Friedrich Witschel, Simone Braun, and Valentin Zacharias. 2010. "Using Machine Learning to Support Continuous Ontology Development." *Knowledge Engineering and Management by the Masses.* Lisbon, Portugal: Springer. pp 381-390.

Russell, S, and P. Norvig. 2010. *Artificial Intelligence: A method for constructing Bayesian Networks.* Third . Pearson Education.

Shaaf, Mohamed. 2000. "Predicting Recession using the Yield Curve: an Artificial Intelligence and Econometric Comparison." *Eastern Economic Journal* 26 (2).

Shapcott, M., R. Sterritt, K. Adamson, and E. Curran. 1999. *NETEXTRACT - Extracting Belief Networks in Telecommunications Data.* http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.6865.

Smets, Frank, and Raf Wouters. 2002. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *International Seminar on Macroeconomics.* European Central Bank.

—. 2007. "Shocks and Frictions in US Business Cycles." *Working Paper Research* (National Bank of Belgium) (109).

Stansbury, Neill, and Catherine Stansbury. 2010. "Engineers against corruption: Preventing corruption in the infrastructure sector." In *Engineering: Issues, Challenges and Opportunities for Development*. UNESCO Report.

Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. 2005. *Introduction to Data Mining.* Pearson.

The European Central Bank. n.d. "Smets-Wouters (2003) Model." Accessed 03 10, 2013. http://wayback.archive.org/web/20130310224052/http://www.ecb.int/home/html/researcher_swm.en.html.

Theodoridis, Sergios, and Konstantinos Koutroumbas. 2008. *Pattern Recognition.* 4th ed. Academic Press.

Trigg, Len, and Eibe Frank. 2015. "Discretize." *Weka doc dev.* Sep 11. http://weka.sourceforge.net/doc.dev/weka/filters/unsupervised/attribute/Discretize.html.

UNESCO. 2010. "Engineering: Issues, Challenges and Opportunities for Development." *UNESCO Report.* http://unesdoc.unesco.org/images/0018/001897/189753e.pdf.

Watermayer, Ron. 2010. "Engineers, Technologists and technicians: How many does a country require?" In *Engineering: Issues, Challenges and Opportunities for Development*. UNESCO Report.

Westholm, Gunnar. 2010. "Engineering indicators - measurements and metrics." In *Engineering: Issues, Challenges and Opportunities for Development*. UNESCO Report.

World Bank. 2015. "“Methodologes.” World Bank Open Data Bank." *World Bank Group.* http://data.worldbank.org/about/data-overview/methodologies.

—. n.d. "Country and Lending Groups." *World Bank data.* Accessed April 16, 2016. http://data.worldbank.org/about/country-and-lending-groups.

—. 2016. "World Bank Open Data Bank." *World Bank Group.* http://databank.worldbank.org.

Zady, M. 2000. "Correlation and Simple Least Squares Regression." *Westgard QC.* August. https://www.westgard.com/lesson42.htm.

APPENDIX

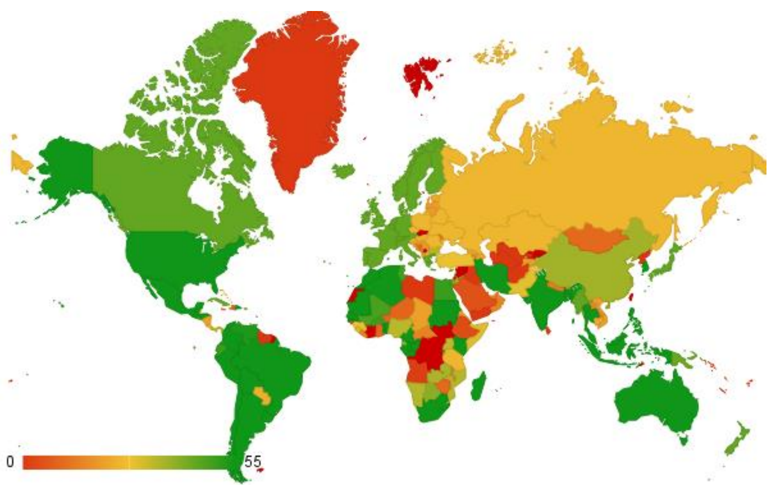
LIST OF APPENDICES

1. GEOGRAPHIC DISTRIBUTION OF DATA AVAILABILITY FOR EACH VARIABLE

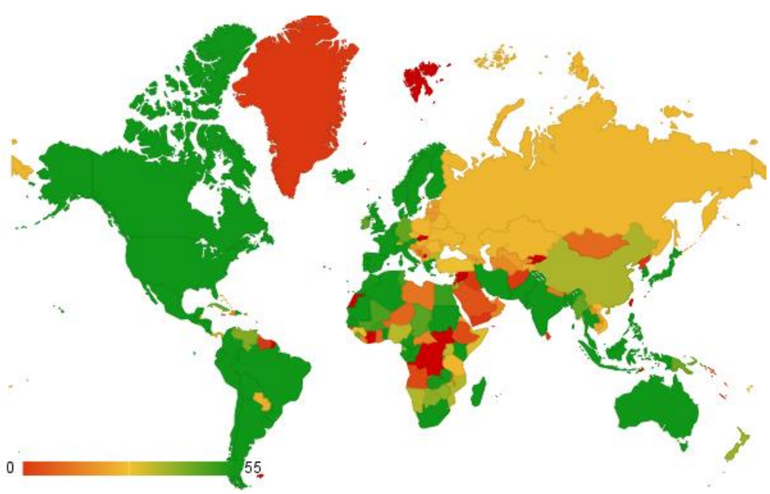
This appendix displays how much data—in the selected variables—is available for each country. Each map shows for the given variable, how much data is available in each country, relative to the other countries. The numbers in the legend are out of 56 and indicate how many years contain data for that country (except for the total, which is out of 1). Variables are ordered alphabetically

Note that Taiwan, French Guiana, Western Sahara, and Svalbard are not considered countries by The World Bank and are therefore marked as having zero data.

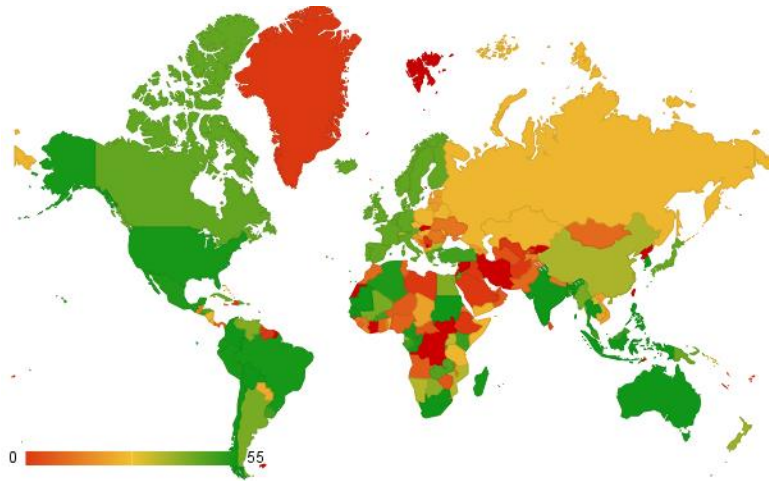
1.1. Availability of data for the variables in Smets and Wouters model



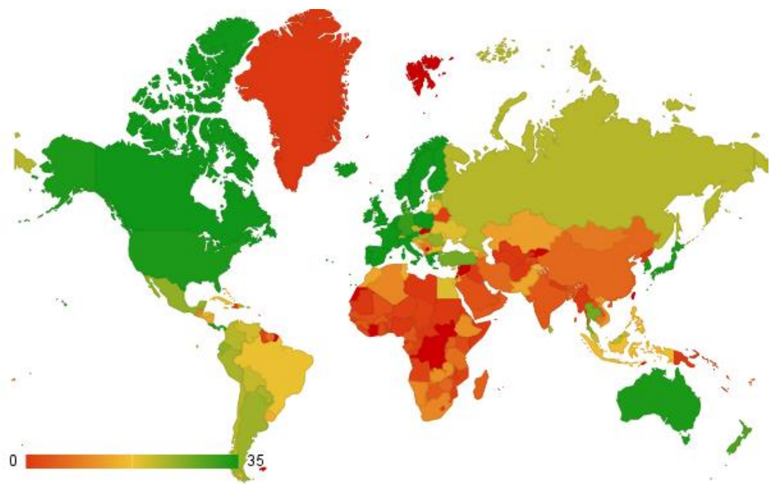
Count of Final consumption expenditure (constant LCU)



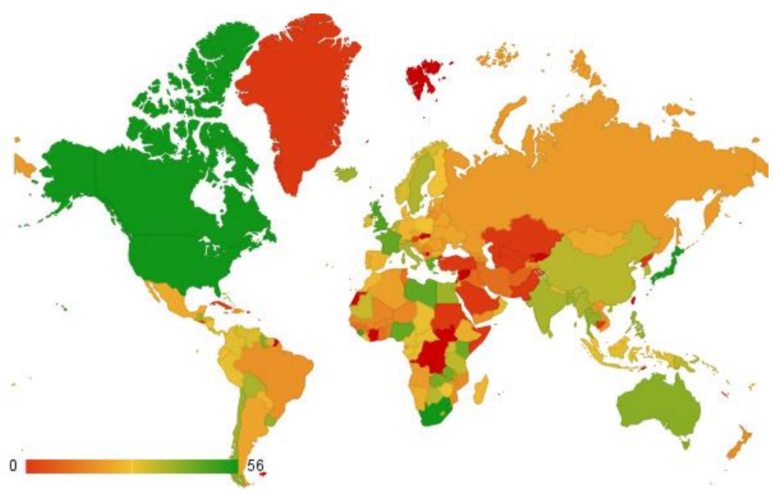
Count of Gross national expenditure (constant LCU)



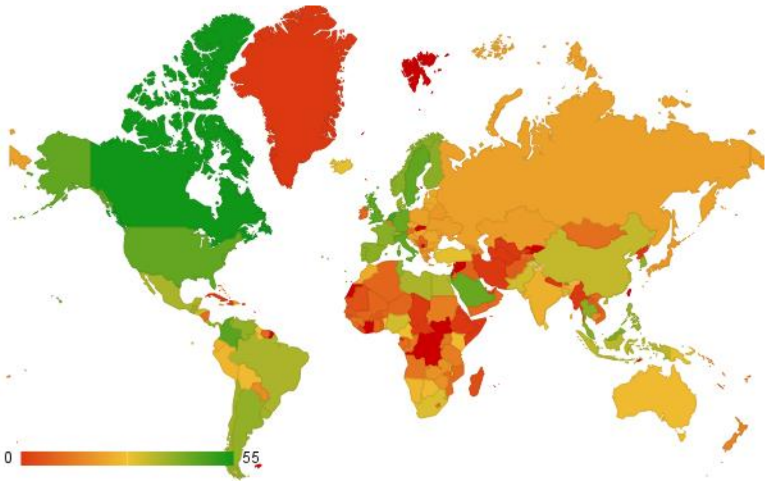
Count of Discrepancy in expenditure estimate of GDP (constant LCU)



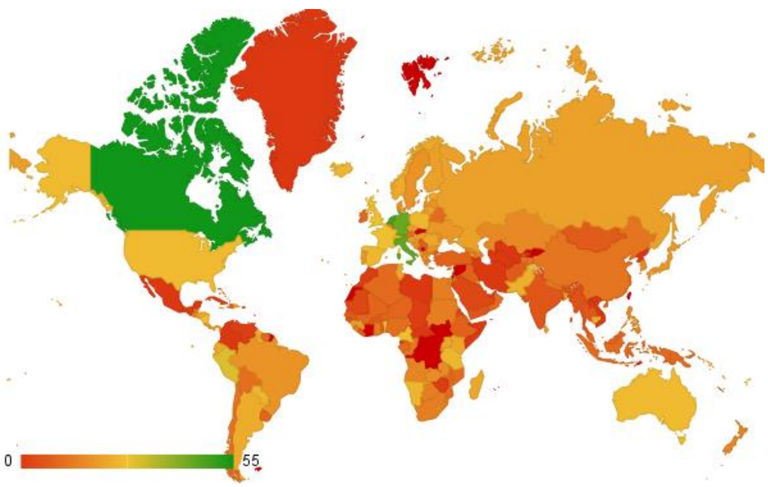
Count of Wage and salaried workers, total (% of total employed)



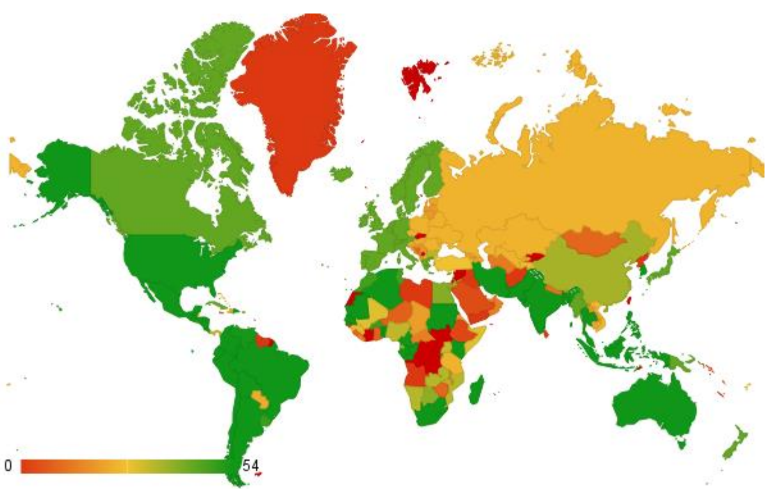
Count of Lending interest rate (%)



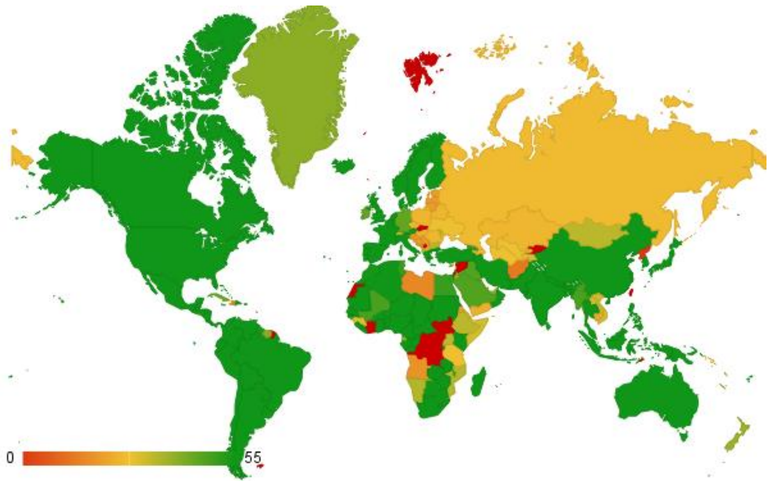
Count of Portfolio Investment, net (BoP, current US$)



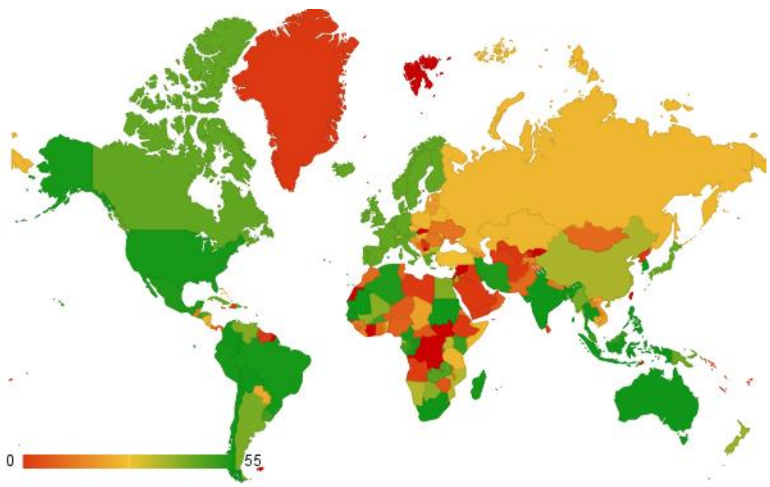
Count of Net capital account (BoP, current US$)



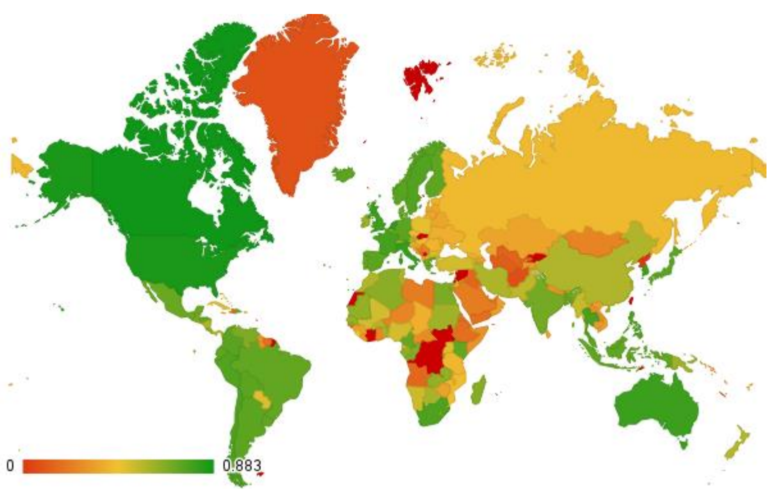
Count of Gross capital formation (annual % growth)



Count of GDP (constant LCU)

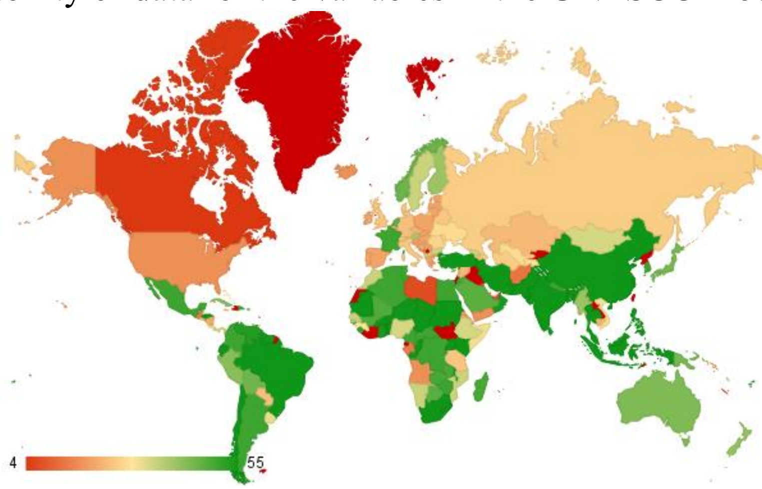


Count of Exogenous expending

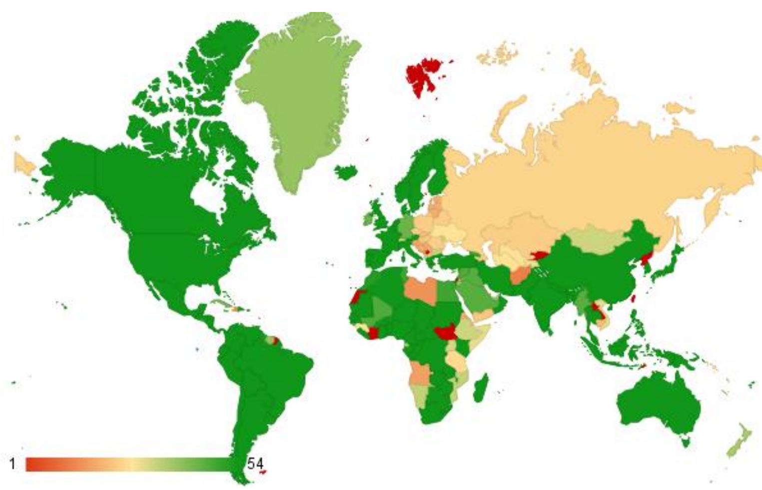


Total ratio

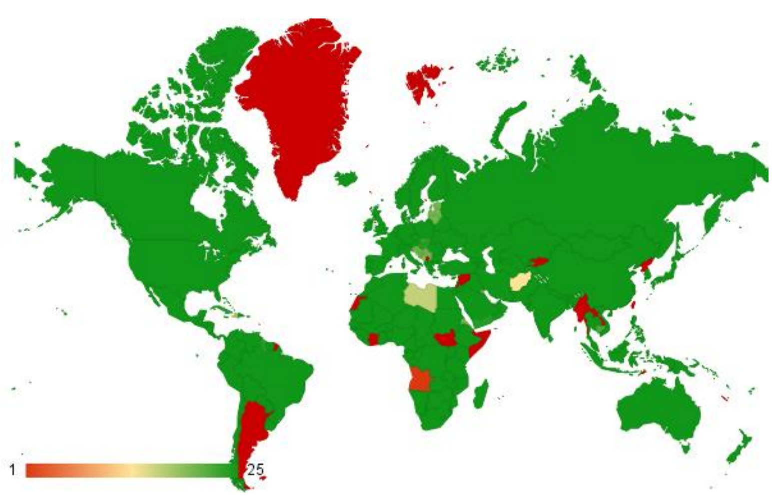
1.2. Availability of data for the variables in the UNESCO model



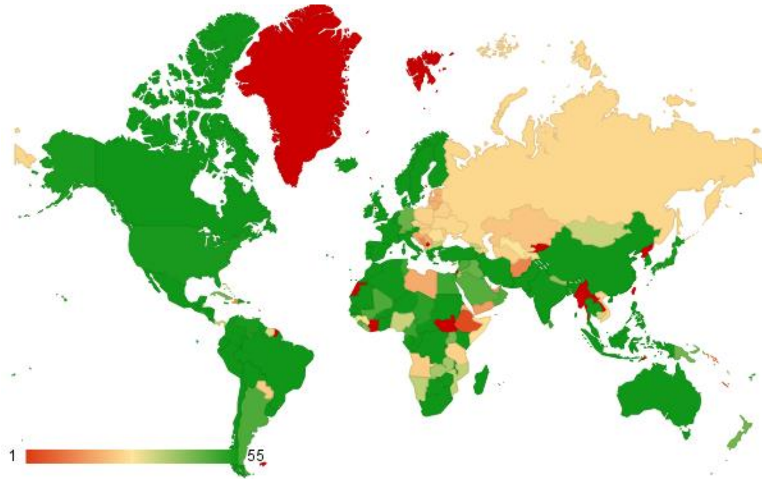
Count of Agriculture, value added (% of GDP)



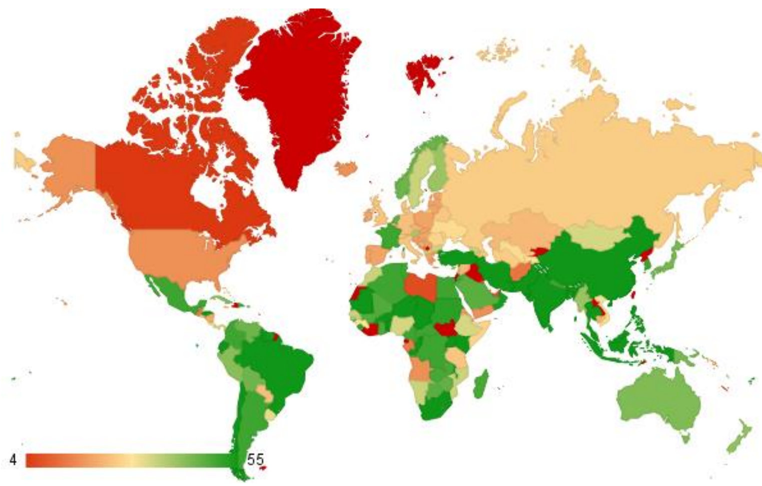
Count of GDP growth (annual)



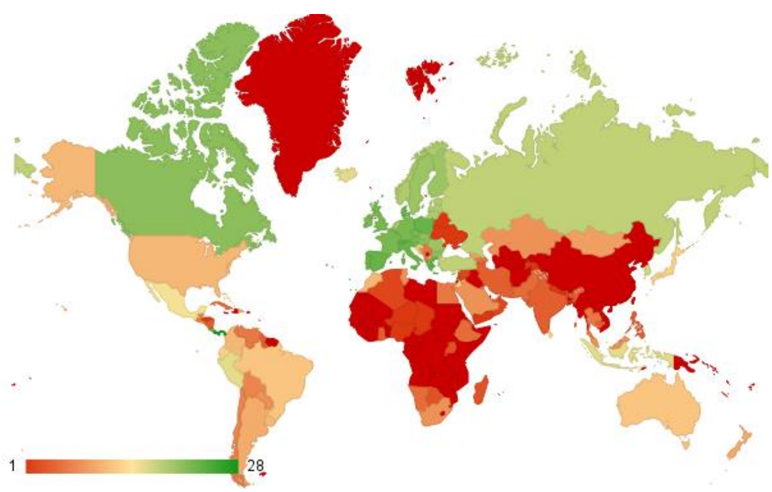
Count of GDP per capita, PPP (constant 2011 international



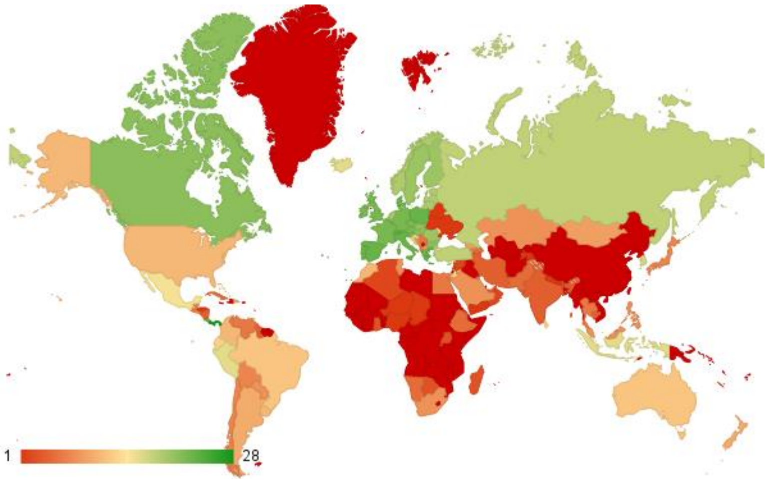
Count of General government final consumption expenditure



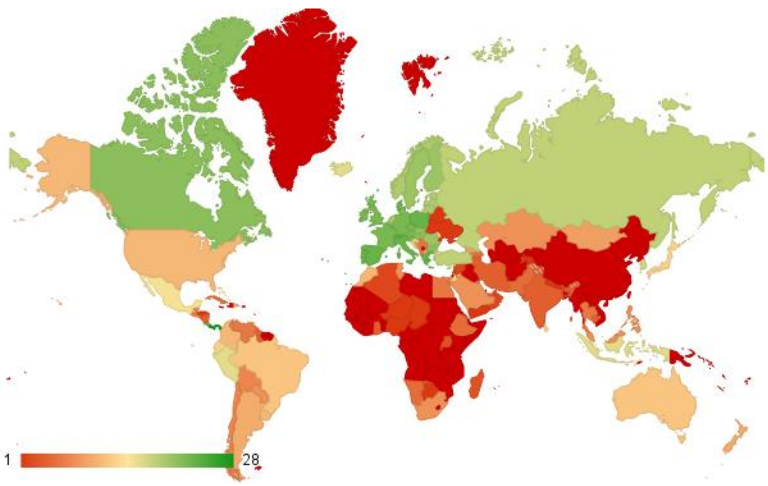
Count of Industry, value added



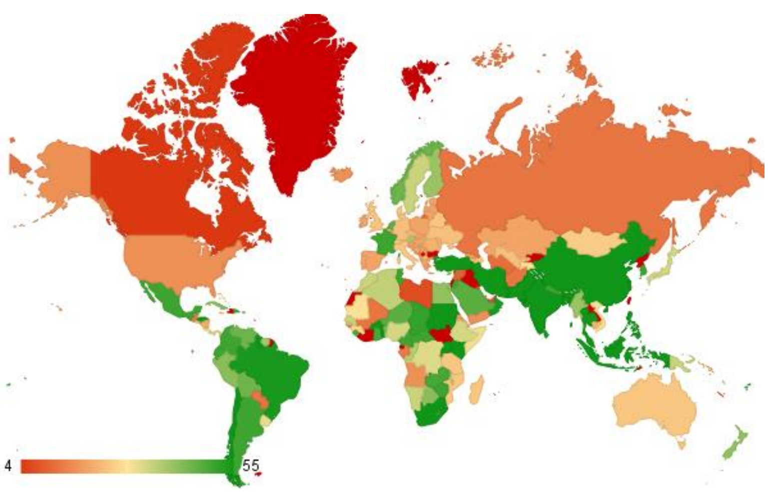
Count of Labor force with primary education



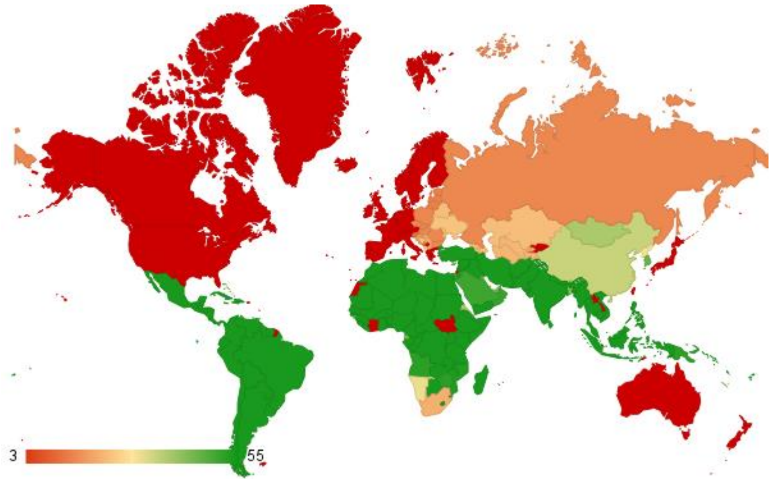
Count of Labor force with secondary education



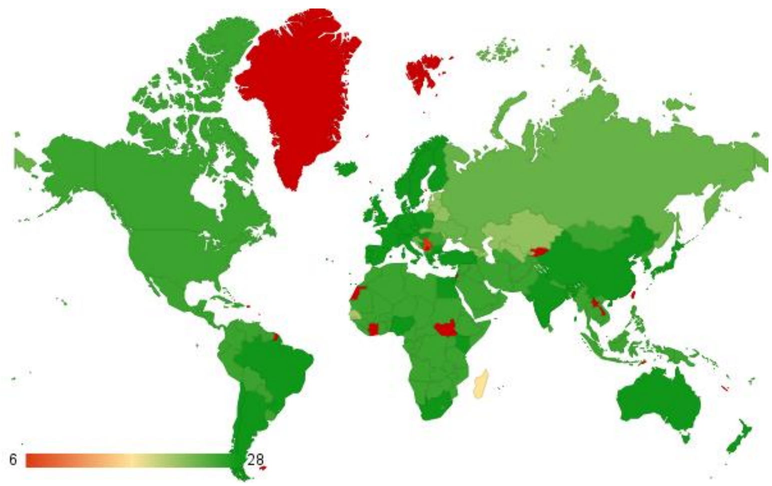
Count of Labor force with tertiary education



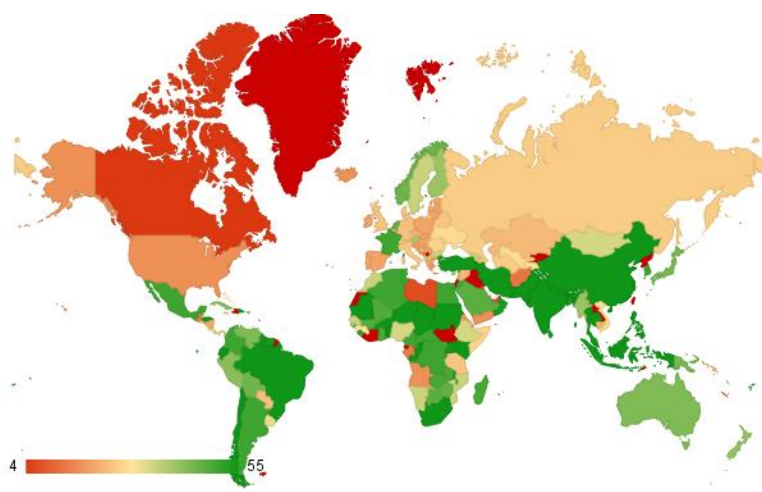
Count of Manufacturing, value added (% of GDP)



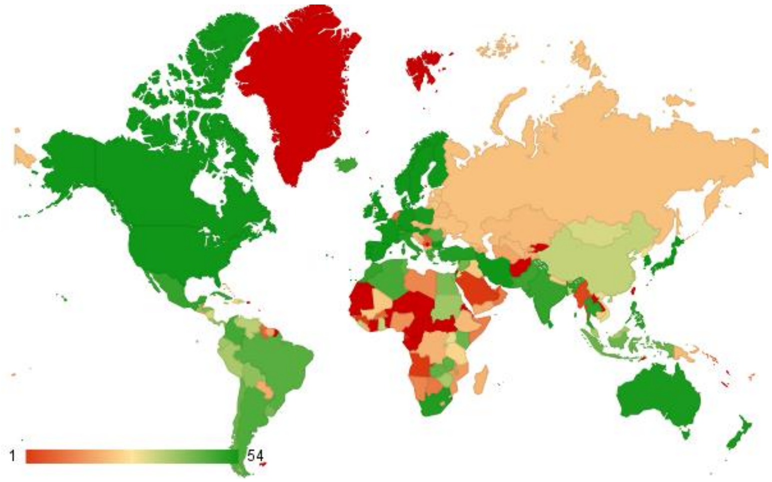
Count of Net official development assistance and official aid received



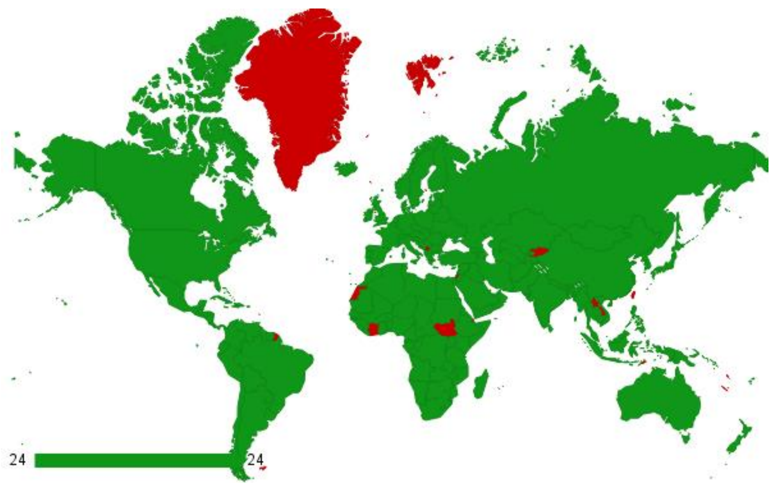
Count of Scientific and technical journal articles



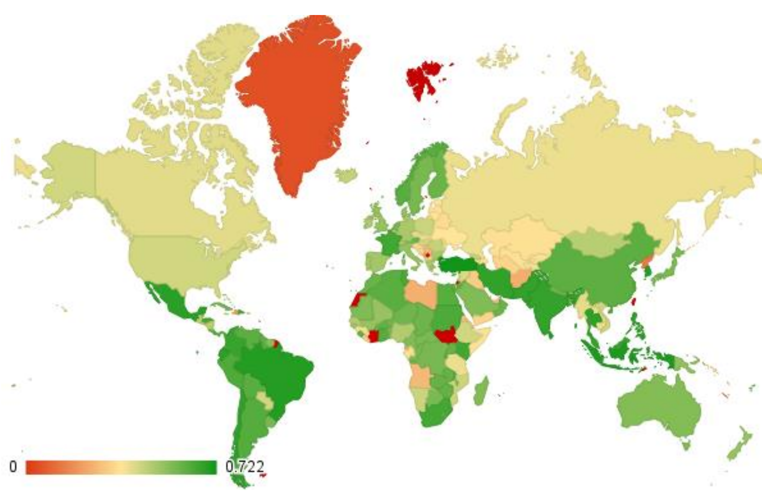
Count of Services, etc., value added (% of GDP)



Count of Trademark applications, total



Count of Unemployment, total (% of total labor force)



Total ratio

2. AVAILABILITY OF DATA FOR DIFFERENT COUNTRIES

This appendix displays how much—of the data for each country, region, or economic group—is available in each of the selected variables. Countries, regions, and economic groups are listed alphabetically, with no distinction (e.g. “Ethiopia” is followed by “Euro area” rather than by the next country, “Faroe Islands”). Numbers are out of 56. A blank cell indicates that the variable has no data for any of the years.

Note that Taiwan, Palestine, French Guiana, and Western Sahara, are not considered countries by The World Bank

2.1. Availability of data for the variables in Smets and Wouters model

Country or Group	Count of Compensation of employees (current LCU)	Count of Inflation, consumer prices (annual %)	Count of Lending interest rate (%)	Count of Exogenous expending	Count of GDP (constant LCU)	Count of Gross capital formation (annual % growth)	Count of Net capital account (BoP, current US$)	Count of Portfolio Investment, net (BoP, current US$)	Count of Wage and salaried workers, total (% of total employed)	Count of Gross national expenditure (constant LCU)	Count of Discrepancy in expenditure estimate of GDP	Count of Final consumption expenditure (constant LCU)
Afghanistan	7	11	9		13		7	7				
Albania	6	24	21	19	35	18	20	15	7	19	19	19
Algeria	18	46	21	53	55	54	10	9	10	55	53	55
American Samoa									3			
Andorra					44							
Angola	14	25	20	1	18		15	12	1	3	7	1
Antigua and Barbuda	13	16	37		38		28	19	2			
Arab World		54										
Argentina	10	22	22	42	55	54	23	39	24	55	42	55
Armenia		31	21	25	25	24	22	19	17	25	25	15
Aruba	20	56	29		16		15	25	6			
Australia	24	56	41	55	55	54	26	26	34	55	55	55
Austria	10	24	2	45	55	44	10	10	35	55	45	45
Azerbaijan		48	16	19	25	21	14	16	11	21	19	21
Bahamas, The	22	49	39	25	55	25	9	16	15	26	25	26
Bahrain	11	29	29	8	35	8	25	38	9	9	28	9
Bangladesh	8	48	39	50	55	54	20	30	6	55	50	55
Barbados	21	23	34		35		5	44	16	30		29
Belarus	19	56	22	25	25	24	10	19	1	25	25	25
Belgium	23	35	29	45	55	44	13	13	35	55	45	45
Belize	13	22	35	34	55	33	19	25	8	35	34	35
Benin			16	23	55	54	26	26	2	55	23	55
Bermuda	20	35			54			8	3			
Bhutan		56	33	15	35	14	9		5	15	15	15
Bolivia	9	9	37	55	55	54	11	26	21	55	55	55

Country or Group	Count of Compensation of employees (current LCU)	Count of Inflation, consumer prices (annual %)	Count of Lending interest rate (%)	Count of Exogenous expending	Count of GDP (constant LCU)	Count of Gross capital formation (annual % growth)	Count of Net capital account (BoP, current US$)	Count of Portfolio Investment, net (BoP, current US$)	Count of Wage and salaried workers, total (% of total employed)	Count of Gross national expenditure (constant LCU)	Count of Discrepancy in expenditure estimate of GDP	Count of Final consumption expenditure (constant LCU)
Bosnia and Herzegovina	14	41	16	9	21	8	17	11	8	9	9	9
Botswana	21	35	35	40	55	39	14	26	4	40	40	40
Brazil		34	18	54	55	54	18	36	18	55	54	55
Brunei Darussalam	23	30	17	25	41	23	9	14	1	25	25	25
Bulgaria	12	55	24	35	35	34	18	21	15	35	35	35
Burkina Faso	6	49	16	50	55	49	6	6	4	50	50	50
Burundi	5	32	34	17	55	17	11		8	18	17	18
Cabo Verde	13	21	41	1	35	4	28	11	1	5	3	5
Cambodia	8	46		22	22	21	23	17	8	22	22	22
Cameroon		56	29	54	55	54	27	27	4	55	54	55
Canada			56	45	55	44	55	55	35	55	45	45
Caribbean small states	6	33							13			
Cayman Islands		31			3				6			
Central African Republic			29	6	55	14				15	6	15
Central Europe and the Baltics	23	56				23			22			
Chad		29	29	21	55	20	1		1	46	21	21
Channel Islands	10	56			10							
Chile		14	38	55	55	54	12	37	22	55	55	55
China	21	50	35	36	55	36	12	33	6	36	36	37
Colombia	18	27	30	55	55	54		46	22	55	55	55
Comoros	24	56	22	32	35	32	10	12		33	32	34
Congo, Dem. Rep.	14	55	9	22	55	21	10	10		55	22	22
Congo, Rep.	22	23	29	54	55	54	13	13	1	55	54	55
Costa Rica			33	23	55	54	17	36	34	55	23	24
Cote d'Ivoire			16	6	55	6	9	9	2	7	6	7
Croatia	24	56	22	20	20	19	19	22	19	20	20	20
Cuba	20	22		44	44	43			17	44	44	44
Curacao	24	56										
Cyprus		22	37	36	40	39	13	27	16	40	36	40
Czech Republic	13	46	22	25	25	24	20	22	22	25	25	25
Denmark	22	56	25	45	55	44	18	38	33	55	45	45
Djibouti	5	56	26	5	25	16	23		1	17	10	
Dominica	19	56	37		38		28	24	4			
Dominican Republic	16	56	24	46	55	54	21	21	23	55	46	55
East Asia & Pacific (all income levels)	3	28				36			6			
East Asia & Pacific (developing only)						36			6			
Ecuador	20	23	27	54	55	54	22	24	25	55	54	50
Egypt, Arab Rep.	21	50	38	41	50	40	14	35	19	50	41	50
El Salvador				49	50	49	22	28	21	50	49	50
Equatorial Guinea	10	46	23	34	35	34	4		1	35	36	35
Eritrea	24	56		12	20	19	2			20	12	20
Estonia	24	56	23	20	20	21	22	22	26	20	20	22
Ethiopia			24		34	3	8		11	4		4

Country or Group	Count of Compensation of employees (current LCU)	Count of Inflation, consumer prices (annual %)	Count of Lending interest rate (%)	Count of Exogenous expending	Count of GDP (constant LCU)	Count of Gross capital formation (annual % growth)	Count of Net capital account (BoP, current US$)	Count of Portfolio Investment, net (BoP, current US$)	Count of Wage and salaried workers, total (% of total employed)	Count of Gross national expenditure (constant LCU)	Count of Discrepancy in expenditure estimate of GDP	Count of Final consumption expenditure (constant LCU)
Euro area		52	7			44	16	16	24			
Europe & Central Asia (all income levels)	1	53				44			23			
Europe & Central Asia (developing only)	16	21				24			12			
European Union	24	24				44			24			
Faroe Islands	15	51										
Fiji	20	56	22		55	25	25	10	5	29		
Finland			28	45	55	44	21	40	35	55	45	45
Fragile and conflict affected situations	19	38										
France			45	45	55	44	27	40	35	55	45	45
French Polynesia	19	56			36		8	9	3			
Gabon	2	10	29	50	55	54	10	11	3	55	50	55
Gambia, The		27	37	10	49	10	15		1	48	10	49
Georgia		19	19	20	50	20	15	17	16	20	20	21
Germany		56	25	45	45	44	44	44	32	45	45	45
Ghana	11	56	11	9	55	8	13	9	2	9	9	9
Greece	8	34	44	45	55	44	16	18	35	55	45	45
Greenland	24	43			40							
Grenada	24	56	37		38		28	25	3			
Guam	23	56							5			
Guatemala	16	56	38	15	55	54	14	38	10	55	15	55
Guinea	20	56	12	12	30	25	20	12	4	27	12	27
Guinea-Bissau		43	11		45	5	9	9		6		6
Guyana	22	56	41		54		23	27	1			
Haiti			20		17	23	9		1	17		10
Heavily indebted poor countries (HIPC)	24	56				24						
High income	24	56				44			23			
High income: nonOECD	22	56				24			22			
High income: OECD		56				44			23			
Honduras	23	46	33	47	55	54	22	32	19	55	47	55
Hong Kong SAR, China	6	21	26	24	50	41	17	17	21	25	24	42
Hungary	20	56	26	24	24	23	20	22	23	24	24	24
Iceland			38	45	55	44	25	29	35	55	45	45
India			37	55	55	54	6	24	4	55	55	55
Indonesia	22	49	30	55	55	54	10	34	17	55	55	55
Iran		12	11	53	55	54			6	55	53	55
Iraq	15	43	10		47		8	8				
Ireland	14	20	31	45	45	44	10	10	35	45	45	45
Isle of Man	7	26			24				1			
Israel	19	24	35		55	19	33	52	19	20		20
Italy	17	5	37	45	55	44	45	45	35	55	45	45
Jamaica	18	39	39	8	49	7	21	19	20	8	8	8
Japan		13	55	44	55	44	19	19	34	55	44	45
Jordan		49	25	38	40	38	9	10	13	39	48	39

Country or Group	Count of Compensation of employees (current LCU)	Count of Inflation, consumer prices (annual %)	Count of Lending interest rate (%)	Count of Exogenous expending	Count of GDP (constant LCU)	Count of Gross capital formation (annual % growth)	Count of Net capital account (BoP, current US$)	Count of Portfolio Investment, net (BoP, current US$)	Count of Wage and salaried workers, total (% of total employed)	Count of Gross national expenditure (constant LCU)	Count of Discrepancy in expenditure estimate of GDP	Count of Final consumption expenditure (constant LCU)
Kazakhstan				25	25	24	16	20	13	25	25	25
Kenya	22	23	44	50	55	54	22	26	2	55	50	55
Kiribati	22	56			45		9	8	3		3	
Korea, Rep.	8	22	35	55	55	54	17	39	34	55	55	55
Kosovo	20	50	11	6	15	8	11	11	1	9	6	9
Kuwait		35	34	5	48	4	22	40	2	5	8	5
Kyrgyz Republic	17	56	19	23	29	24	21	20	10	23	25	25
Lao PDR	22	31	19	14	31	18		6	3	17	31	15
Latin America & Caribbean (all income levels)	14	26				54			17			
Latin America & Caribbean (developing only)	24	56				54			17			
Latvia			21	20	20	19	18	23	19	20	20	20
Least developed countries: UN classification		29				22						
Lebanon	24	52	32	20	27	20	10	10	2	21	20	21
Lesotho	11	56	35	54	55	53	20	12	2	54	54	54
Liberia			29	15	55	14	8	4	2	15	15	15
Libya	17	21	46		16			36		12		
Liechtenstein					40							
Lithuania	19	23	18	20	20	19	21	22	17	20	20	20
Low & middle income		10				36			1			
Low income	20	56				21						
Lower middle income	3	28				54			3			
Luxembourg		55	19	45	55	44	13	13	35	55	45	45
Macao SAR, China	22	13	28	32	33	32	13	13	18	33	32	33
Macedonia, FYR	4	51	21	25	25	24	16	19	19	25	25	25
Madagascar	24	56	26	54	55	54	22	4	5	55	54	55
Malawi			35	8	55	11	12	25	1	12	8	12
Malaysia	13	56	28	50	55	54	9	41	21	55	50	55
Maldives	23	43	19	4	14	3	8	8	5	4	4	4
Mali		51	16	40	48	29	9	9	3	48	42	48
Malta	10	56	38	41	44	40	21	44	15	41	41	41
Marshall Islands					34		9	9	1		2	
Mauritania	24	56	33	55	55	54	3	5		55	55	55
Mauritius	23	15	34	39	39	38	1	24	12	39	39	39
Mexico	15	56	23	53	55	54		36	25	55	53	55
Micronesia, Fed. Sts.			20		29		5	5				
Middle East & North Africa (all income levels)	12	56				13			2			
Middle East & North Africa (developing only)	13	43				48			3			
Middle income	8	56				36			1			
Moldova	23	56	20	23	35	23	18	21	16	23	23	10
Monaco	21	56			39							
Mongolia	20	45	23	10	34	9	7	11	8	10	10	10
Montenegro	24	56	9	14	18	14	7	8	7	15	14	15
Morocco			26	10	54	48	1	23	14	54	10	55

Country or Group	Count of Compensation of employees (current LCU)	Count of Inflation, consumer prices (annual %)	Count of Lending interest rate (%)	Count of Exogenous expending	Count of GDP (constant LCU)	Count of Gross capital formation (annual % growth)	Count of Net capital account (BoP, current US$)	Count of Portfolio Investment, net (BoP, current US$)	Count of Wage and salaried workers, total (% of total employed)	Count of Gross national expenditure (constant LCU)	Count of Discrepancy in expenditure estimate of GDP	Count of Final consumption expenditure (constant LCU)
Mozambique	7	36	17	35	35	34	10	10	1	35	35	35
Myanmar	24	25	36	42	48	44	5			45	42	45
Namibia	15	23	24	35	35	34	25	25	10	35	35	35
Nepal	9	47	31	13	55	12	13		1	14	13	14
Netherlands	3	53	35	45	55	44	48	48	35	55	45	45
New Caledonia	13	11			36		9	9	1			
New Zealand	11	19	17	38	38	44	15	15	30	38	38	45
Nicaragua		52	28	25	55	54	27	11	14	55	25	21
Niger	10	48	16	7	55	8	9	9	1	9	7	9
Nigeria	6	21	45	8	55	33	11	31		34	8	34
North America	20	45				54			23			
North Korea	17	27										
Northern Mariana Islands		55										
Norway	23	55	31	45	55	44	23	40	35	55	45	45
Not classified												
OECD members	17	22				44			23			
Oman	21	23	31	7	55	14	10	10	4	15	7	15
Other small states		43				34						
Pacific island small states												
Pakistan	24	56		10	55	54	24	31	14	55	10	29
Palau		6			24							
Panama	24	56	29	9	55	32	20	38	32	33	9	34
Papua New Guinea	23	56	35	44	55	44	10	29		44	44	44
Paraguay	17	35	25	23	55	23	25	17	23	24	23	24
Peru	13	49	29	54	55	54	30	25	23	55	54	55
Philippines			39	55	55	54	16	38	16	55	55	55
Poland		41	28	25	25	24	25	21	35	25	25	25
Portugal	2	55	24	45	55	44	19	40	35	55	45	45
Puerto Rico	12	56		41	54	41			22	43	41	42
Qatar		49	26		15	11	4	4	9	15	25	15
Romania	24	56	21	13	25	24	22	22	25	25	13	25
Russian Federation	22	56	20	25	26	24	21	21	22	25	25	25
Rwanda	12	53	31	54	55	54	18	19	3	55	54	55
Samoa	6	14	13		33		12	11	5	12		
San Marino	4	50	22		39				15			
Sao Tome and Principe	24	56	26		15		18	5	1			
Saudi Arabia		13			47	4	3	44	3	5		5
Senegal	10	48	16	55	55	54	7	7	3	55	55	55
Serbia		39	18		20	19	8	8	10	20		20
Seychelles	13	56	26	20	55	19	22	26	2	20	29	20
Sierra Leone	23	32	50	10	55	9	13	11	1	44	10	31
Singapore	15	56	38	40	55	39	8	43	24	40	40	40
Sint Maarten (Dutch part)												

Country or Group	Count of Compensation of employees (current LCU)	Count of Inflation, consumer prices (annual %)	Count of Lending interest rate (%)	Count of Exogenous expending	Count of GDP (constant LCU)	Count of Gross capital formation (annual % growth)	Count of Net capital account (BoP, current US$)	Count of Portfolio Investment, net (BoP, current US$)	Count of Wage and salaried workers, total (% of total employed)	Count of Gross national expenditure (constant LCU)	Count of Discrepancy in expenditure estimate of GDP	Count of Final consumption expenditure (constant LCU)
Slovak Republic			16	23	23	22	21	22	21	23	23	23
Slovenia			19	20	20	24	22	23	22	20	20	25
Small states	14	35				25						
Solomon Islands	14	23	34		25		22	12			29	
Somalia	6	7		25	31	29				30	25	30
South Africa	24	27	56	55	55	54	14	30	10	55	55	55
South Asia	25	56				54			4			
South Sudan	23	56		6	7	6				7	6	7
Spain			25	45	55	44	25	40	35	55	45	45
Sri Lanka	9	38	37		54		20	28	24		2	
St. Kitts and Nevis	16	55	34		38		28	22	2			
St. Lucia		20	37		35		28	24	8			
St. Martin (French part)												
St. Vincent and the Grenadines	9	18	38		55		28	22	4			
Sub-Saharan Africa (all income levels)		23				34						
Sub-Saharan Africa (developing only)	17	30				38						
Sudan	8	43		55	55	54	14	14		55	55	55
Suriname		41	24		40		10	10	7			
Swaziland		48	44	9	45	31	18	33	1	42	9	32
Sweden		23	36	45	55	44	18	45	35	55	45	45
Switzerland		28	34	35	35	44	24	38	24	35	35	45
Syrian Arab Republic		28	33	29	48	32	4	3	7	34	29	29
Tajikistan		56	18	15	30	28	12	10	2	24	15	21
Tanzania		23	34	25	27	24	25	14	7	25	25	25
Thailand		22	40	55	55	54	6	40	27	55	55	55
Timor-Leste		56	13	11	16	13	9	9	1	14	11	14
Togo		26	16	16	55	54	9	9	1	55	16	55
Tonga		42	35		34		13	8	2			
Trinidad and Tobago		36	35	49	55	54		17	22	55	49	55
Tunisia		22	12	9	50	48	22	39	15	49	9	49
Turkey		56		27	55	27	8	29	27	28	49	28
Turkmenistan		50		1	28	13				16	5	
Turks and Caicos Islands		56							6			
Tuvalu		34			25		9	9	1			
Uganda		40	32	33	33	32	22	13	5	33	33	33
Ukraine		34	22	12	28	24	19	20	19	25	12	25
United Arab Emirates		42		2	40	13			3	14	2	14
United Kingdom		46	48	45	55	44	28	45	35	55	45	45
United States		46	56	54	55	54	26	45	34	55	54	55
Upper middle income		41				36			6			
Uruguay		56	39	49	55	49	9	37	15	55	49	55
Uzbekistan					28	24				20		
Vanuatu		56	34	11	36	10	33	17	1	11	11	11

Country or Group	Count of Compensation of employees (current LCU)	Count of Inflation, consumer prices (annual %)	Count of Lending interest rate (%)	Count of Exogenous expending	Count of GDP (constant LCU)	Count of Gross capital formation (annual % growth)	Count of Net capital account (BoP, current US$)	Count of Portfolio Investment, net (BoP, current US$)	Count of Wage and salaried workers, total (% of total employed)	Count of Gross national expenditure (constant LCU)	Count of Discrepancy in expenditure estimate of GDP	Count of Final consumption expenditure (constant LCU)
Venezuela, RB		36	31	41	55	53	1	39	20	41	41	53
Vietnam		38	20	21	31	25		10	12	26	21	21
Virgin Islands (U.S.)		38			20							
West Bank and Gaza		51	14	21	21	20	20	20	18	21	21	21
World		44				44			1			
Yemen, Rep.		46	18		24		6	10	4		25	
Zambia		40	43	47	55	34	17	18	11	54	47	36
Zimbabwe		39	28	6	55	9	1	18	6	39	6	10

2.2. Availability of data for the variables in the UNESCO model

Country or group	Count of Labor force with primary education (% of total)	Count of GDP per capita, PPP (constant 2011 international $)	Count of GDP growth (annual %)	Count of Unemployment, total (% of total labor force)	Count of Services, etc., value added (% of GDP)	Count of Manufacturing, value added (% of GDP)	Count of Agriculture, value added (% of GDP)	Count of General government final consumption expenditure	Count of Labor force with secondary education (% of total)	Count of Labor force with tertiary education (% of total)	Count of Scientific and technical journal articles	Count of Trademark applications, total	Count of Net official development assistance and official aid received	Count of Industry, value added (% of GDP)
Afghanistan		13	12	24	13	13	13	13			26		54	13
Albania	6	25	34	24	35	19	35	35	6	6	26	20	26	35
Algeria	2	25	54	24	50	36	50	55	2	2	26	48	54	50
American Samoa														

Country or group	Count of Labor force with primary education (% of total)	Count of GDP per capita, PPP (constant 2011 international $)	Count of GDP growth (annual %)	Count of Unemployment, total (% of total labor force)	Count of Services, etc., value added (% of GDP)	Count of Manufacturing, value added (% of GDP)	Count of Agriculture, value added (% of GDP)	Count of General government final consumption expenditure	Count of Labor force with secondary education (% of total)	Count of Labor force with tertiary education (% of total)	Count of Scientific and technical journal articles	Count of Trademark applications, total	Count of Net official development assistance and official aid received	Count of Industry, value added (% of GDP)
Andorra			43		14	14	14				26	17		14
Angola		1	17	24	17	17	17	24			26	1	52	17
Antigua and Barbuda	1	25	37		38	38	38	38	1	1	26	14	41	38
Arab World		25	39	24	16	10	16	47			26	33	55	16
Argentina	10		54	24	50	50	50	48	10	10	28	47	54	50
Armenia	7	25	24	24	25	25	25	25	7	7	22	20	23	25
Aruba	1	1	15		18	18	18	18	1	1		3	20	18
Australia	12	25	54	24	43	25	43	55	12	12	28	53		43
Austria	22	25	54	24	39	39	39	55	22	22	28	54		39
Azerbaijan	5	25	24	24	25	25	25	25	5	5	22	19	23	25
Bahamas, The	12	25	54	24	30	26	30	37	12	12	26	20	43	26
Bahrain	4	25	34	24	16	16	16	35	4	4	26	35	45	16
Bangladesh		25	54	24	55	55	55	53			26	34	43	55
Barbados	2	25	34	24				35	2	2	26	21	45	
Belarus	1	25	24	24	25	25	25	25	1	1	22	21	23	25
Belgium	23	25	54	24	20	20	20	55	23	23	28	53		20
Belize	9	25	54	24	36	36	36	35	9	9	26	13	54	36
Benin		25	54	24	55	44	55	55			26		54	55
Bermuda	5	24	53		17	12	17	5	5	5			39	17
Bhutan		25	34	24	35	35	35	35			26	17	51	35
Bolivia	7	25	54	24	45	45	45	55	7	7	26	40	54	45
Bosnia and Herzegovina	7	21	20	24	21	21	21	13	7	7	28	18	25	21
Botswana	3	25	54	24	47	42	47	55	3	1	26	11	54	47
Brazil	12	25	54	24	55	54	55	55	12	12	28	46	54	55
Brunei Darussalam		25	40	24	40	40	40	36			26	25	45	40
Bulgaria	21	25	34	24	35		35	35	21	21	28	54	15	35
Burkina Faso		25	54	24	55	55	55	55			26	8	54	55
Burundi		25	54	24	45	45	45	55			26	31	54	45
Cabo Verde		25	34	24	35	21	35	5			26	1	43	35
Cambodia	4	22	21	24	22	22	22	33	4	4	26	24	54	22
Cameroon		25	54	24	50	50	50	50			26		54	50
Canada	21	25	54	24	4	4	4	55	21	21	26	54		4
Caribbean small states		25	48	24	29	25	29	38			26	7	55	25
Cayman Islands	6	1	1			7	7		6	6			31	7
Central African Republic		25	54	24	50	50	50	55			26		54	50
Central Europe and the Baltics	21	25	23	24	20	20	20	25	21	21	24	46	55	20
Chad	1	25	54	24	55	51	55	52	1	1	26		54	55
Channel Islands			9											
Chile	7	25	54	24	55	55	55	55	7	7	28	44	54	55
China		25	54	24	55	55	55	55			28	34	35	55
Colombia	11	25	54	24	50	50	50	55	11	11	26	48	54	50
Comoros		25	34	24	35	35	35	35			26	1	48	35
Congo, Dem. Rep.		25	54	24	50	33	50	55			26	20	54	50
Congo, Rep.		25	54	24	55	37	55	55			26		54	55
Costa Rica	23	25	54	24	31	31	31	55	23	23	26	38	54	31

Country or group	Count of Labor force with primary education (% of total)	Count of GDP per capita, PPP (constant 2011 international $)	Count of GDP growth (annual %)	Count of Unemployment, total (% of total labor force)	Count of Services, etc., value added (% of GDP)	Count of Manufacturing, value added (% of GDP)	Count of Agriculture, value added (% of GDP)	Count of General government final consumption expenditure	Count of Labor force with secondary education (% of total)	Count of Labor force with tertiary education (% of total)	Count of Scientific and technical journal articles	Count of Trademark applications, total	Count of Net official development assistance and official aid received	Count of Industry, value added (% of GDP)
Cote d'Ivoire		25	54	24	55	40	55	55			26		54	55
Croatia	16	20	19	24	20	20	20	20	16	16	24	22	19	20
Cuba	5	24	43	24	42	42	42	44	5	5	26	50	54	42
Curacao												4		
Cyprus	17	25	39	24	40	40	40	40	17	17	26	42	45	40
Czech Republic	22	25	24	24	22	22	22	25	22	22	28	21	15	22
Denmark	23	25	54	24	25	25	25	55	23	23	28	54		25
Djibouti		25	24		19	19	19	16			26	1	48	19
Dominica	1	25	37		38	38	38	38	1	1	26	22	41	38
Dominican Republic	10	25	54	24	55	50	50	55	10	10	26	33	54	50
East Asia & Pacific (all income levels)		25	54	24	44	34	44	54			28	34	55	44
East Asia & Pacific (developing only)		25	54	24	55	55	55	55			28	34	55	55
Ecuador	13	25	54	24	55	50	55	55	13	13	26	38	54	55
Egypt, Arab Rep.	6	25	49	24	50	41	50	50	6	6	28	42	54	50
El Salvador	8	25	49	24	25	25	25	50	8	8	26	17	54	25
Equatorial Guinea		25	34	24				51			26	1	41	
Eritrea		20	19	24	18	18	18	20			26		38	18
Estonia	19	20	19	24	20	20	20	20	19	19	22	22	14	20
Ethiopia	5	25	33	24	34	34	34	4	5	5	26	21	54	34
Euro area	23	25	54	24	24	24	24	55	23	23	28	54	55	24
Europe & Central Asia (all income levels)	16	25	54	24	24	24	24	55	16	16	28	22	55	24
Europe & Central Asia (developing only)		25	27	24	30	24	30	28			26	21	55	30
European Union	23	25	54	24	24	24	24	55	23	23	28	54	55	24
Faeroe Islands					16	16		16						
Fiji		25	54	24	52	52	52	51			26	13	54	52
Finland	20	25	54	24	40	40	40	55	20	20	28	54		40
Fragile and conflict affected situations		25	24	24	12	5	14	19			26		55	12
France	22	25	54	24	50	50	50	55	22	22	28	54		50
French Polynesia	3		35				11	10	3	3			39	
Gabon		25	54	24	14	14	14	55			26		54	14
Gambia, The		24	47	24	10	10	10	37			26	21	54	10
Georgia	9	25	49	24	35	19	35	35	9	9	22	22	23	35
Germany	22	25	44	24	24	24	24	45	22	22	28	54		24
Ghana	4	25	54	24	49	49	54	55	3	4	26	34	54	49
Greece	23	25	54	24	20	20	20	55	23	23	28	48		20
Greenland			39											
Grenada		25	37		38	38	38	38			26	2	41	38
Guam														
Guatemala	6	25	54	24	14	14	14	55	6	6	26	39	54	14
Guinea		25	28	24	29	27	29	29			26	1	54	29
Guinea-Bissau		25	44	24	45	15	45	45			26	8	46	45
Guyana	4	24	53	24	55	55	55	55	4	4	26	8	54	55
Haiti		17	16	24				16			26	27	54	
Heavily indebted poor countries (HIPC)		25	54	24	45	31	45	55			26		55	44
High income	16	25	54	24	17	17	17	54	10	16	26	54	55	17

Country or group	Count of Labor force with primary education (% of total)	Count of GDP per capita, PPP (constant 2011 international $)	Count of GDP growth (annual %)	Count of Unemployment, total (% of total labor force)	Count of Services, etc., value added (% of GDP)	Count of Manufacturing, value added (% of GDP)	Count of Agriculture, value added (% of GDP)	Count of General government final consumption expenditure	Count of Labor force with secondary education (% of total)	Count of Labor force with tertiary education (% of total)	Count of Scientific and technical journal articles	Count of Trademark applications, total	Count of Net official development assistance and official aid received	Count of Industry, value added (% of GDP)
High income: nonOECD	11	25	25	24	26	14	26	26	11	11	20	22	55	26
High income: OECD	16	25	54	24	17	17	17	54	11	16	26	54	55	17
Honduras	1	25	54	24	55	55	55	55	1	1	26	42	54	55
Hong Kong SAR, China	11	25	49	24	15	15	15	50	11	11	11	32	45	15
Hungary	21	24	23	24	20	20	20	24	21	21	28	54	15	20
Iceland	16	25	54	24	17	17	17	55	16	16	28	49		17
India	4	25	54	24	55	55	55	55	4	4	28	50	54	55
Indonesia	16	25	54	24	55	55	55	55	16	16	26	44	54	55
Iran, Islamic Rep.	4	25	54	24	55	55	55	55	4	4	26	54	54	55
Iraq		25	46	24				45			26	23	54	
Ireland	22	25	44	24	20	20	20	55	22	22	28	54		20
Isle of Man			23											
Israel	12	25	54	24				55	12	12	28	54	45	
Italy	23	25	54	24	25	25	25	55	23	23	28	52		25
Jamaica		25	48	24	22	22	22	54			26	39	53	22
Japan	13	25	54	24	44	34	44	54	7	13	28	54		44
Jordan	13	25	39	24	50	50	50	39	13	13	26	31	54	50
Kazakhstan	8	25	24	24	23	20	23	23	8	8	22	21	23	23
Kenya		25	54	24	55	55	55	55			28	44	54	55
Kiribati		25	44		36	36	36	21			26	1	54	36
Korea, Dem. Rep.				24							26	29	30	
Korea, Rep.	17	25	54	24	50	50	50	55	17	17	28	54	45	50
Kosovo	1	15	14		9	9	9	10	1	1	15		5	9
Kuwait	5	20	46	24	5	5	5	50	5	5	26	7	44	5
Kyrgyz Republic	5	25	28	24	25	25	27	28	5	5	26	21	22	25
Lao PDR		25	30	24	26	26	26	20			26	22	54	26
Latin America & Caribbean (all income levels)	7	25	54	24	50	49	50	55	7	7	28	46	55	50
Latin America & Caribbean (developing only)	8	25	54	24	50	49	50	55	8	8	28	43	55	50
Latvia	19	20	19	24	20	20	20	20	19	19	22	22	14	20
Least developed countries: UN classification		25	33	24	35	31	35	30			26	1	55	33
Lebanon	1	25	26	24	21	21	21	26	1	1	26	25	54	21
Lesotho		25	54	24	52	51	55	54			26	19	54	52
Liberia		25	54	24				42			26	20	54	
Libya		16	15	24	7	7	7	19			26	13	54	7
Liechtenstein			39								26	49		15
Lithuania	19	20	19	24	20	20	20	20	19	19	22	23	14	20
Low & middle income		25	54	24	55	55	55	55			28	33	55	55
Low income		25	32	24	26	25	29	25			26	1	55	26
Lower middle income		25	54	24	55	55	55	55			28	49	55	55
Luxembourg	22	25	54	24	20	20	20	55	22	22	28	11		20
Macao SAR, China	17	25	32	24	22	22	22	33	17	17		24	39	22
Macedonia, FYR	9	25	24	24	25	25	25	25	9	9	26	21	21	25

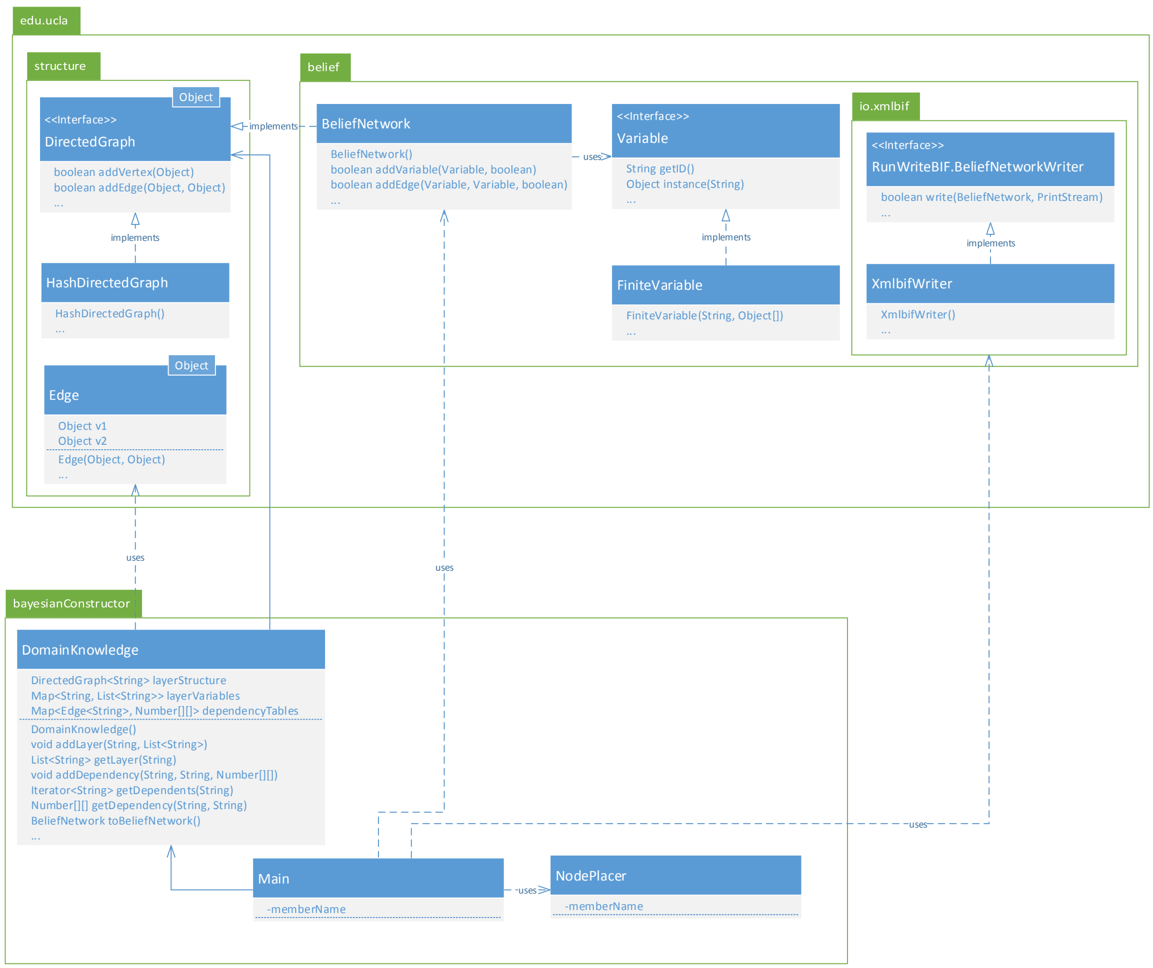
Country or group	Count of Labor force with primary education (% of total)	Count of GDP per capita, PPP (constant 2011 international $)	Count of GDP growth (annual %)	Count of Unemployment, total (% of total labor force)	Count of Services, etc., value added (% of GDP)	Count of Manufacturing, value added (% of GDP)	Count of Agriculture, value added (% of GDP)	Count of General government final consumption expenditure	Count of Labor force with secondary education (% of total)	Count of Labor force with tertiary education (% of total)	Count of Scientific and technical journal articles	Count of Trademark applications, total	Count of Net official development assistance and official aid received	Count of Industry, value added (% of GDP)
Madagascar	3	25	54	24	49	25	49	55	3	3	17	20	54	49
Malawi		25	54	24	55	40	55	55			26	40	54	55
Malaysia	8	25	54	24	55	55	55	55	8	8	26	37	54	55
Maldives		14	13	24	20	20	20	11			26	6	54	20
Mali		25	47	24	48	13	48	48			26	23	54	48
Malta	15	24	43	24	41	41	41	42	15	15	26	49	45	41
Marshall Islands		25	33		9	9	9				26		23	9
Mauritania		25	54	24	55	30	55	55			26		54	55
Mauritius	6	25	38	24	39	39	39	39	6	6	26	30	54	39
Mexico	15	25	54	24	50	50	50	55	15	15	26	50	54	50
Micronesia, Fed. Sts.		25	28		20	21	21	1			26		23	20
Middle East & North Africa (all income levels)		25	46	24	12	7	12	47			26	39	55	12
Middle East & North Africa (developing only)		25	49	24	50	40	50	50			26	42	55	50
Middle income		25	54	24	55	55	55	55			28	33	55	55
Moldova	14	25	34	24	26	22	26	24	14	14	22	21	22	26
Monaco			38								26	53		
Mongolia	9	25	33	24	34	26	34	34	9	9	26	32	38	34
Montenegro	7	18	17	24	15	15	15	15	7	7	6	7	11	15
Morocco	11	25	52	24	35	35	35	55	11	11	26	50	54	35
Mozambique		25	34	24	35	25	35	35			26	16	53	35
Myanmar			46	24	40	40	40				26	3	54	40
Namibia	5	25	34	24	35	35	35	35	5	5	26	13	31	35
Nepal	1	25	54	24	55	50	50	40	1	1	26	24	54	50
Netherlands	19	25	54	24	46	46	46	55	19	19	28	11		46
New Caledonia	1		35		8	8	8	7	1	1			39	8
New Zealand	10	25	37	24	41	41	41	44	10	10	28	53		41
Nicaragua	4	25	54	24	21	21	21	55	4	4	26	32	54	21
Niger	1	25	54	24	55	47	55	55	1	1	26		54	55
Nigeria	1	25	54	24	34	34	34	34	1	1	28	15	54	34
North America	11	25	54	24	17	17	17	54	11	11	26	54	55	17
Northern Mariana Islands													41	
Norway	19	25	54	24	45	45	45	55	19	19	28	54		45
Not classified														
OECD members	17	25	54	24	17	17	17	54	11	17	26	54	55	17
Oman	2	25	54	24	54	54	54	48	2	2	26	11	51	54
Other small states		25	34	24	26	26	26	37			22	4	55	26
Pacific island small states		25	35		36	36	36	31			26	1	55	36
Pakistan	5	25	54	24	55	55	55	55	5	5	26	49	54	55
Palau		24	23		23	23	23				17		22	23
Panama	28	25	54	24	33	33	33	33	28	28	26	39	54	33
Papua New Guinea		25	54	24	44	35	44	44			26	21	52	44
Paraguay	8	25	54	24	24	13	24	24	8	8	26	19	54	24
Peru	16	25	54	24	42	42	42	55	16	16	26	37	54	42
Philippines	4	25	54	24	55	55	55	55	8	8	26	51	54	55

Country or group	Count of Labor force with primary education (% of total)	Count of GDP per capita, PPP (constant 2011 international $)	Count of GDP growth (annual %)	Count of Unemployment, total (% of total labor force)	Count of Services, etc., value added (% of GDP)	Count of Manufacturing, value added (% of GDP)	Count of Agriculture, value added (% of GDP)	Count of General government final consumption expenditure	Count of Labor force with secondary education (% of total)	Count of Labor force with tertiary education (% of total)	Count of Scientific and technical journal articles	Count of Trademark applications, total	Count of Net official development assistance and official aid received	Count of Industry, value added (% of GDP)
Poland	23	25	24	24	20	20	20	25	23	23	28	53	15	20
Portugal	23	25	54	24	20	20	20	55	23	23	28	54		20
Puerto Rico		24	53	24	43	43	43	54						43
Qatar		15	14	24	15	15	15	21			26	12	39	15
Romania	20	25	24	24	27	22	27	27	20	20	26	46	15	27
Russian Federation	18	25	25	24	26	13	26	26	18	18	24	22	15	26
Rwanda	2	25	54	24	50	50	50	55	2	2	26	36	54	50
Samoa		25	32								26	21	53	
San Marino	10		38						10	10	26	18		
Sao Tome and Principe		15	14		14	14	14				26	9	44	14
Saudi Arabia	8	25	46	24	47	47	47	47	8	8	26	1	50	47
Senegal		25	54	24	35	35	35	55			21		54	35
Serbia	6	20	19	24	20	19	20	20	6	6	6	8	20	20
Seychelles		25	54		39	39	39	39			26	21	54	39
Sierra Leone		25	54	24	51	49	51	47			26	34	54	51
Singapore	20	25	54	24	40	40	40	55	20	20	28	40	45	40
Sint Maarten (Dutch part)		1										4		
Slovak Republic	20	23	22	24	20	20	20	25	20	20	23	21	15	20
Slovenia	22	20	19	24	20	20	20	20	22	22	24	22	12	20
Small states		25	38	24	30	26	30	38			26		55	26
Solomon Islands		25	24	24	17	17	17	10			26	7	54	17
Somalia			30	24	27	31	31	27			26	12	54	27
South Africa	8	25	54	24	55	55	55	55	8	8	28	52	21	55
South Asia	4	25	54	24	55	55	55	55	4	4	28	50	55	55
South Sudan		7	6					7					3	
Spain	23	25	54	24	20	20	20	55	23	23	28	54		20
Sri Lanka	13	25	53	24	55	55	55	51	13	13	26	54	54	55
St. Kitts and Nevis	1	25	37		38	38	38	38	1	1	26		41	38
St. Lucia	8	25	34		36	36	36	35	8	8	26	12	41	36
St. Martin (French part)														
St. Vincent and the Grenadines	2	25	54		38	38	38	38	2	2	26	2	41	38
Sub-Saharan Africa (all income levels)		25	54	24	50	39	50	55			28		55	49
Sub-Saharan Africa (developing only)		25	54	24	50	39	50	55			28		55	49
Sudan		25	54	24	55	55	55	55			26	38	54	52
Suriname		25	39	24	54	39	54	31			26	18	54	54
Swaziland		25	44	24	50	50	55	52			26	29	54	50
Sweden	20	25	54	24	35	35	35	55	20	20	28	54		35
Switzerland	24	25	34	24	25	25	25	45	24	24	28	50		25
Syrian Arab Republic	2		47	24	23	10	23	48	2	2	26	43	54	23
Tajikistan	1	25	29	24	29	29	29	29	1	1	26	20	22	29
Tanzania		25	26	24	25	25	25	25			26	25	54	25
Thailand	6	25	54	24	55	55	55	55	6	6	26	51	54	55
Timor-Leste	1	16	15	24	14	14	14	14	1	1			37	14
Togo		25	54	24	55	55	55	54			26		54	55
Tonga		25	33		40	40	40	38			26	11	54	40

Country or group	Count of Labor force with primary education (% of total)	Count of GDP per capita, PPP (constant 2011 international $)	Count of GDP growth (annual %)	Count of Unemployment, total (% of total labor force)	Count of Services, etc., value added (% of GDP)	Count of Manufacturing, value added (% of GDP)	Count of Agriculture, value added (% of GDP)	Count of General government final consumption expenditure	Count of Labor force with secondary education (% of total)	Count of Labor force with tertiary education (% of total)	Count of Scientific and technical journal articles	Count of Trademark applications, total	Count of Net official development assistance and official aid received	Count of Industry, value added (% of GDP)
Trinidad and Tobago	15	25	54	24	30	30	30	55	15	15	26	30	50	30
Tunisia	8	25	49	24	50	50	50	50	8	8	26	45	54	50
Turkey	18	25	54	24	55	55	55	55	18	18	28	52	54	55
Turkmenistan		25	27	24	26	12	26	25			26	18	22	26
Turks and Caicos Islands													34	
Tuvalu		25	24		14	13	14				17	1	39	14
Uganda	3	25	32	24	55	55	55	43	3	3	26	30	54	55
Ukraine	1	25	27	24	28	24	28	26	1	1	24	22	24	28
United Arab Emirates	3	25	39	24	7	7	7	14	3	3	26	3	45	7
United Kingdom	22	25	54	24	25	25	25	55	22	22	28	54		25
United States	11	25	54	24	17	17	17	54	11	11	26	54		17
Upper middle income		25	54	24	55	55	55	55			28	34	55	55
Uruguay	11	25	54	24	32	32	32	55	11	11	26	43	54	32
Uzbekistan		25	27	24	28	24	28	28			22	21	22	28
Vanuatu		25	35		36	36	36	35			26		54	36
Venezuela, RB	6	25	54	24	43	45	53	54	6	6	26	35	54	45
Vietnam		25	30	24	30	30	30	26			26	30	54	30
Virgin Islands (U.S.)			19											
West Bank and Gaza	9	21	20	24	20	20	20	21	9	9			21	20
World		25	54	24	19	17	19	54			28	34	55	19
Yemen, Rep.	1	24	23	24	17	17	17	17	1	1	26	19	54	17
Zambia		25	54	24	49	49	49	38			26	45	54	49
Zimbabwe		25	54	24	46	50	50	55			26	37	52	46

3. CLASS DIAGRAMS

Java notation is used when appropriate



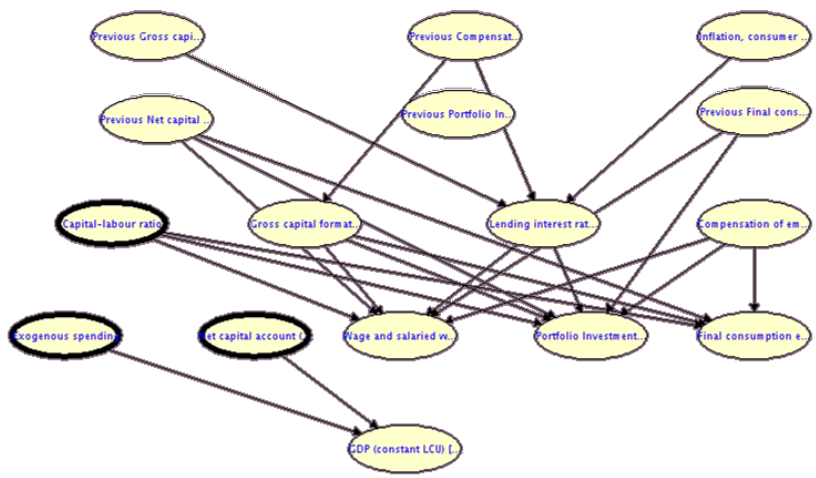
Class diagram for the Bayesian Constructor. DomainKnowledge stores the graph and table representing the domain knowledge model, as well as the subgraphs to generate the Bayesian network. Main runs the Bayesian network constructor. NodePlacer computes the pixel coordinates for each node so that they will be readable when seen with an XML-BIF viewer.

4. GENERATED BAYESIAN NETWORKS

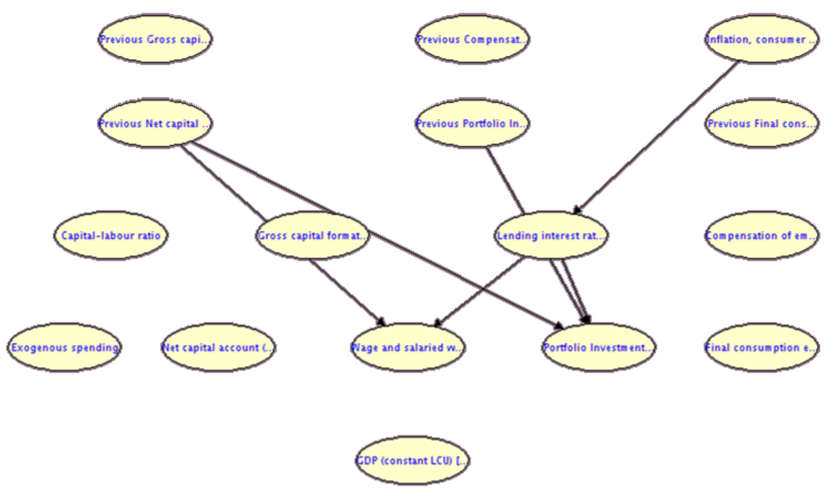
This appendix contains a sample of the Bayesian networks generated by the program for each of the countries

4.1. Networks generated using the Smets and Wouters Domain Knowledge model

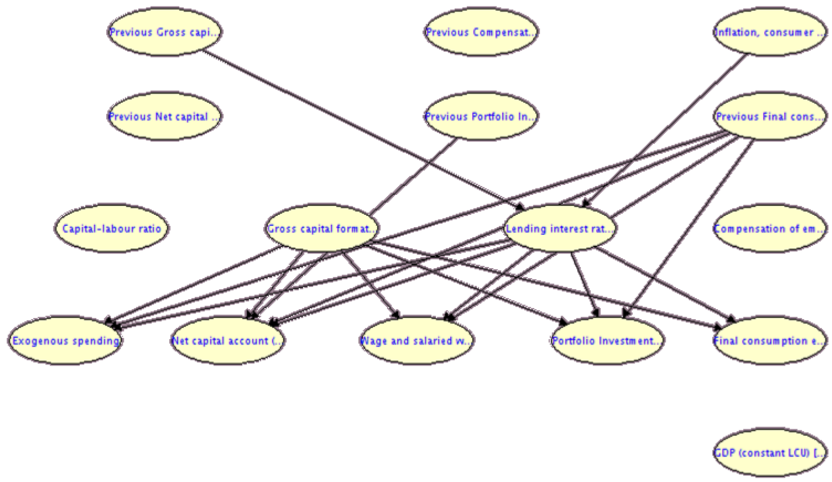
Eleven representative countries and regions are shown in alphabetical order'



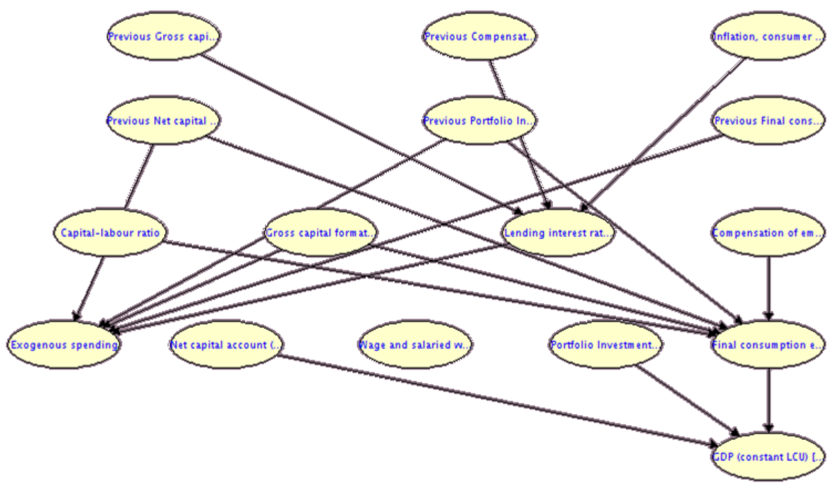
USA



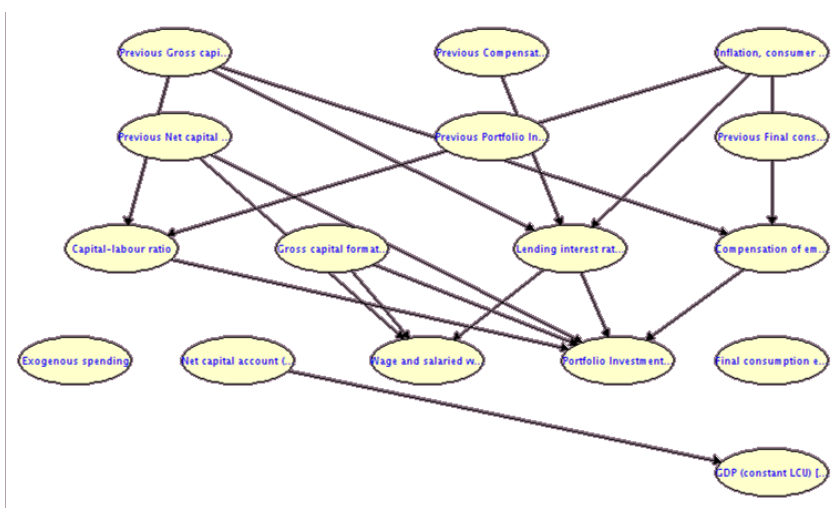
Euro Area



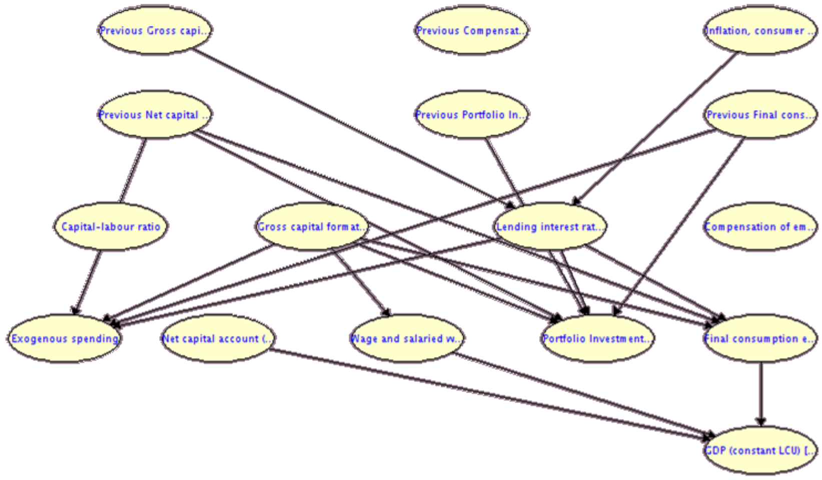
Canada



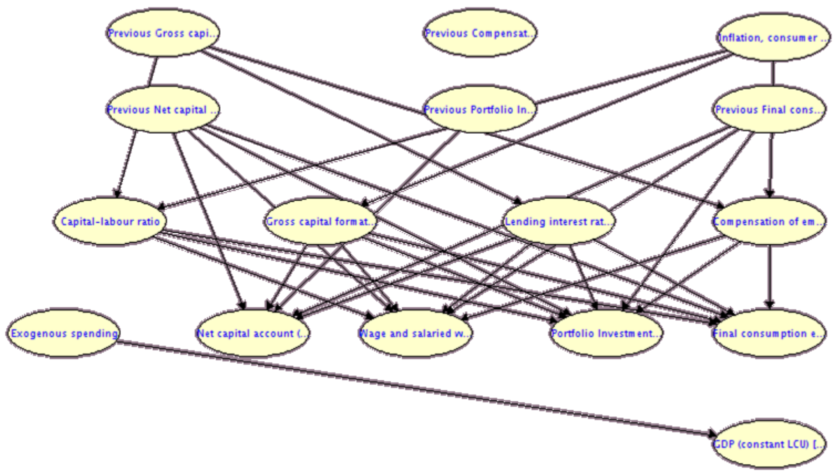
Congo, Rep.



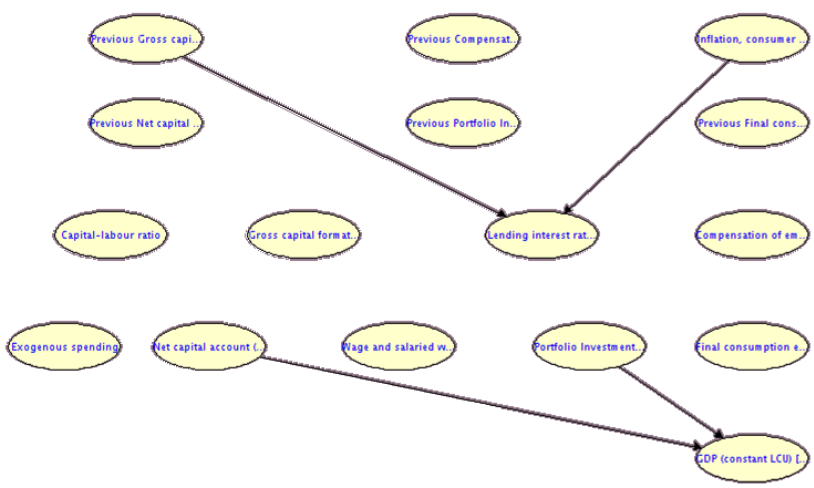
Fiji



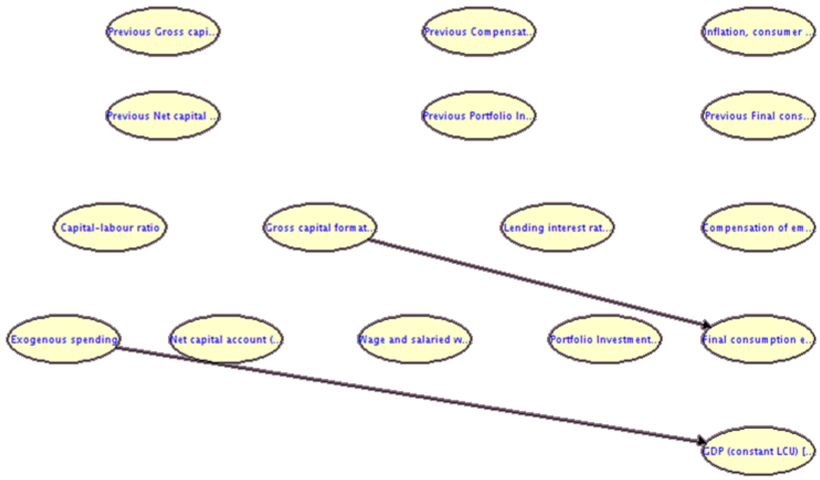
Gabon



United Kingdom

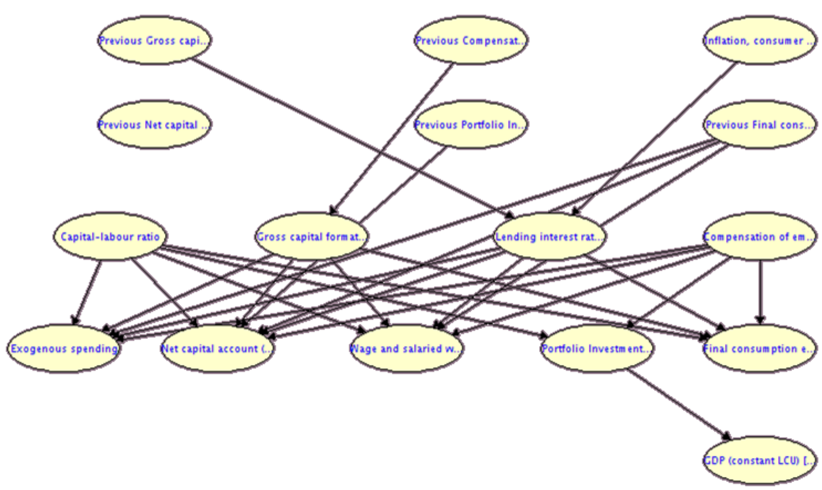


Iraq



Somalia

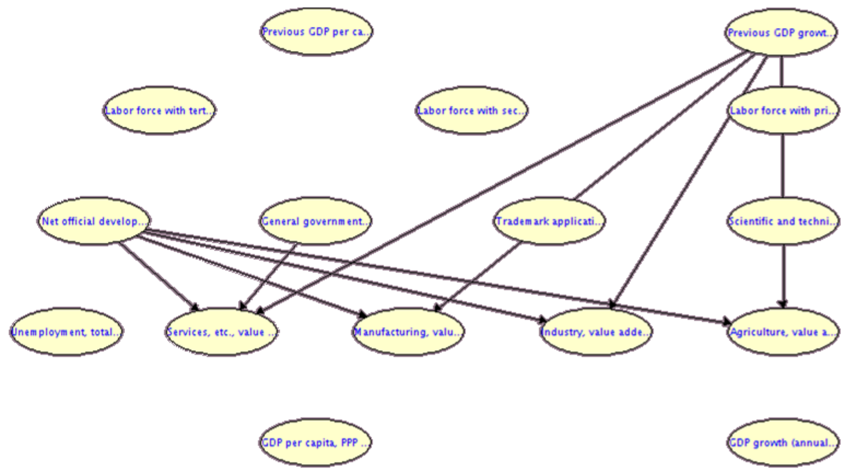
Sub-Saharan Africa (all income levels)



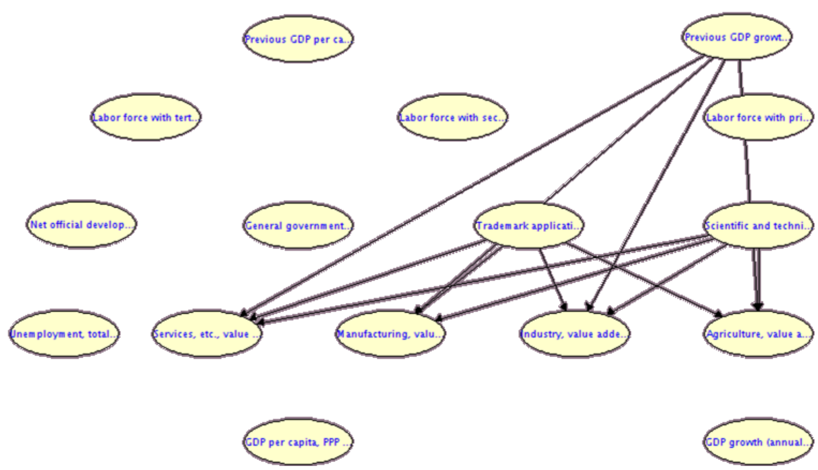
Sweden

4.2. Networks generated using the UNESCO Domain Knowledge model

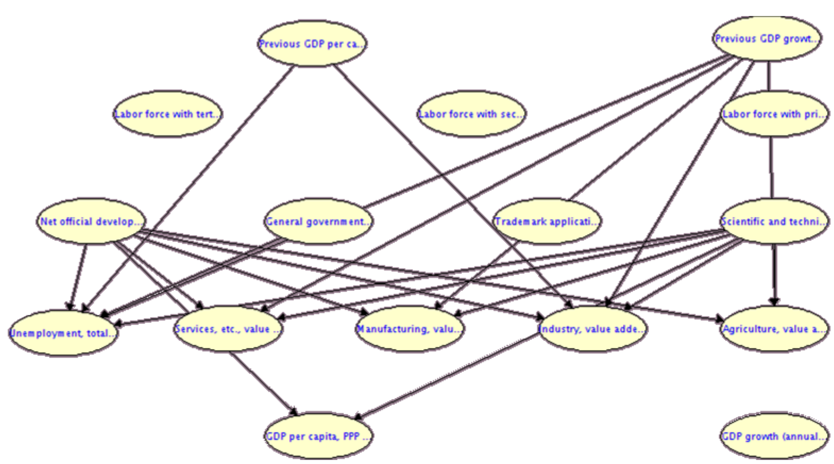
The first 56 countries and regions are shown ordered by their country code.



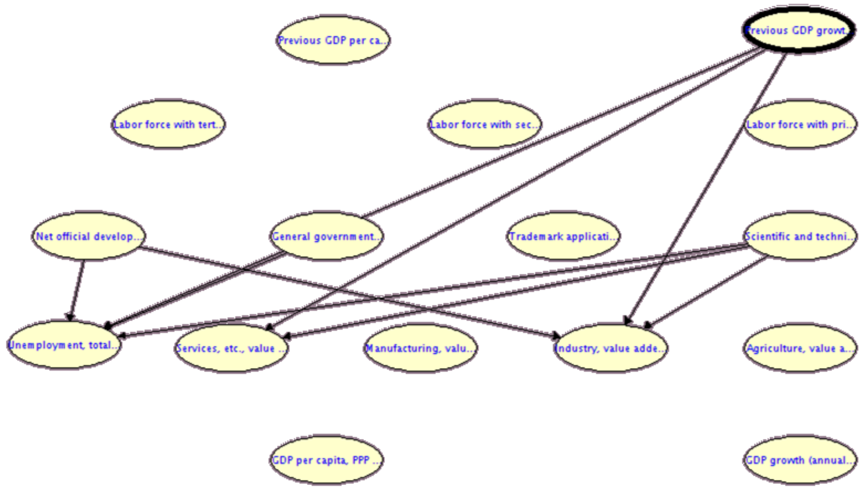
Aruba



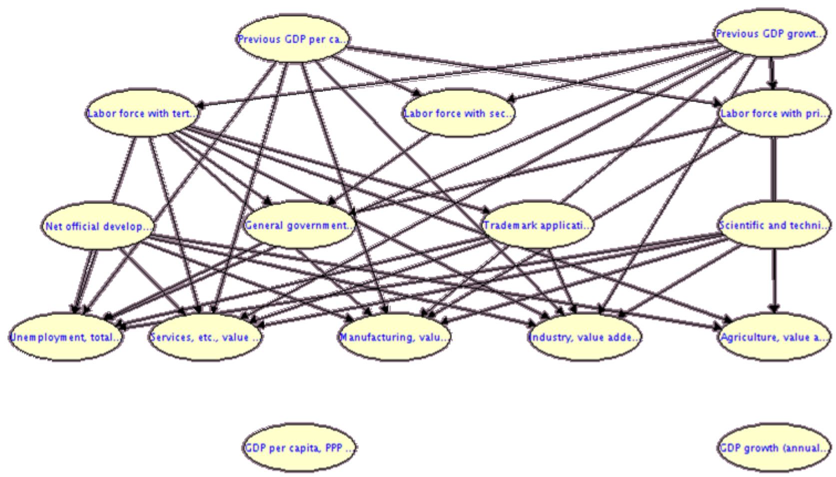
Andorra



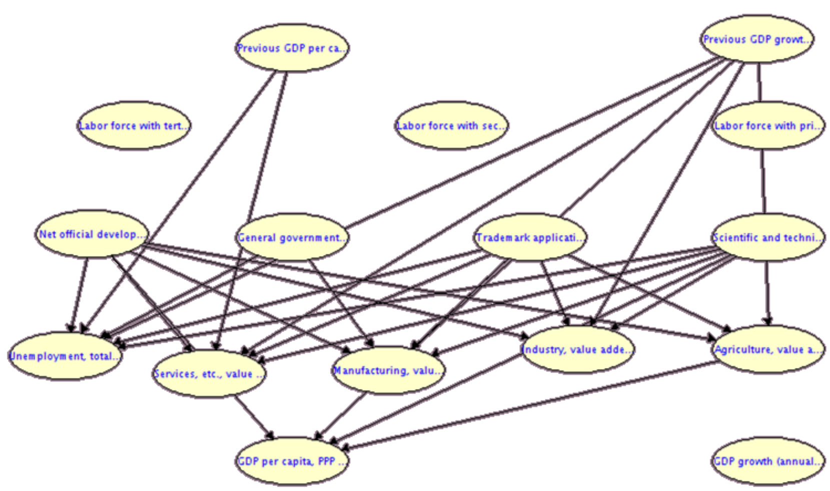
Afghanistan



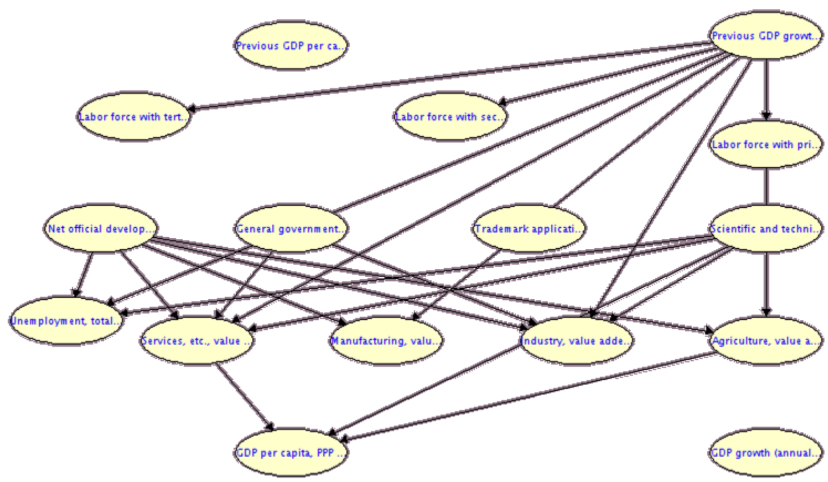
Angola



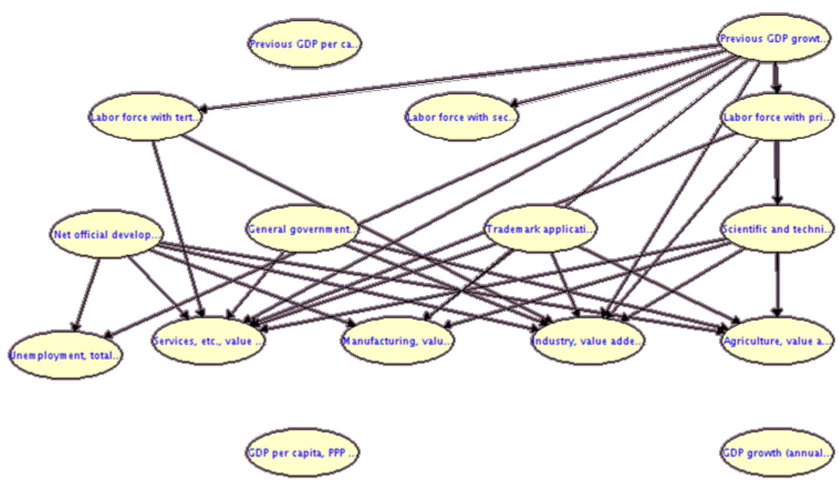
Albania



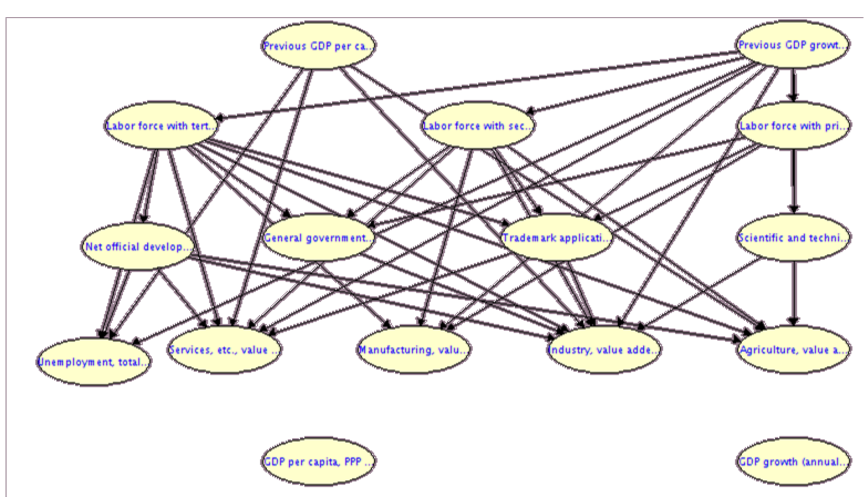
Arab World



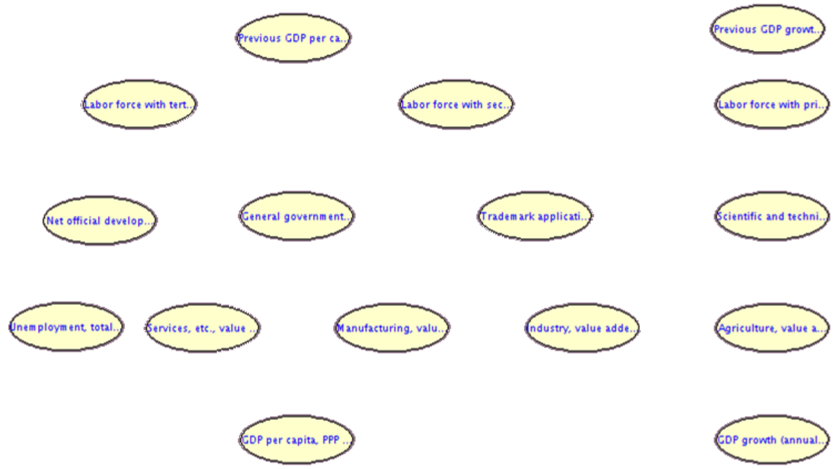
United Arab Emirates



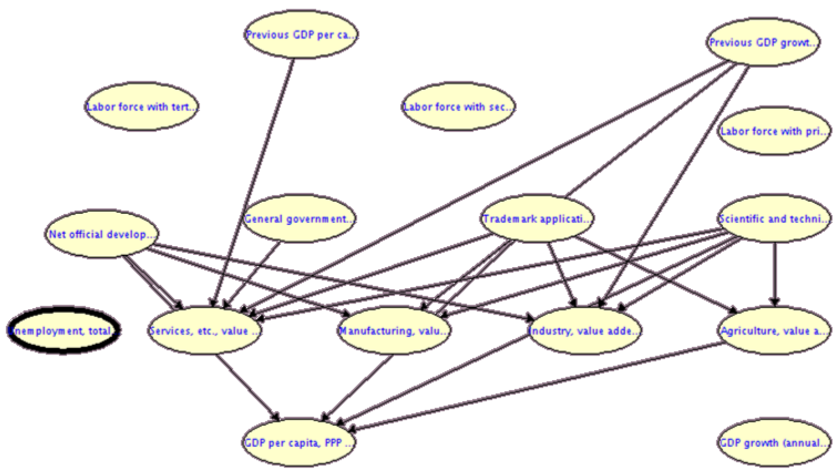
Argentina



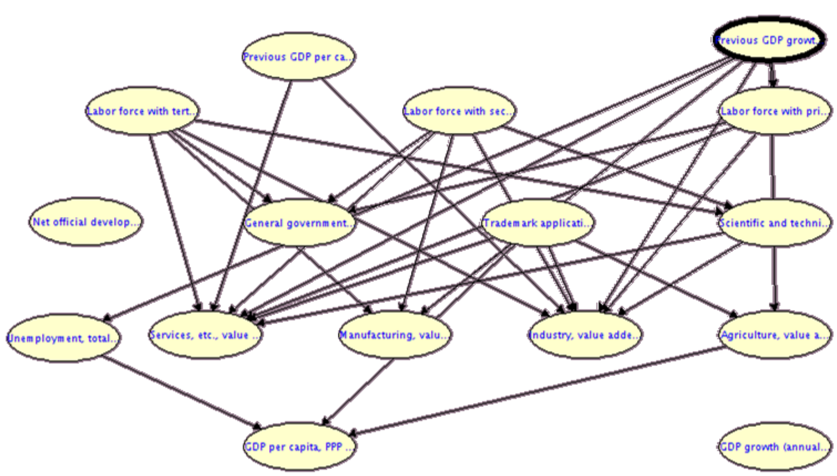
Armenia



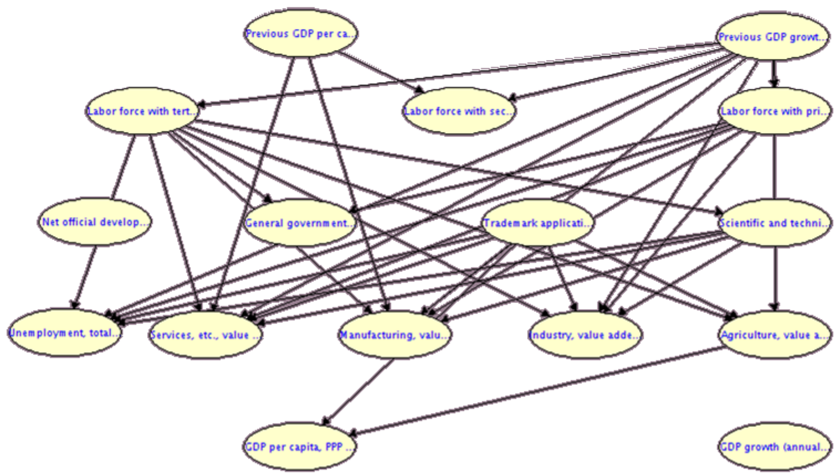
American Samoa



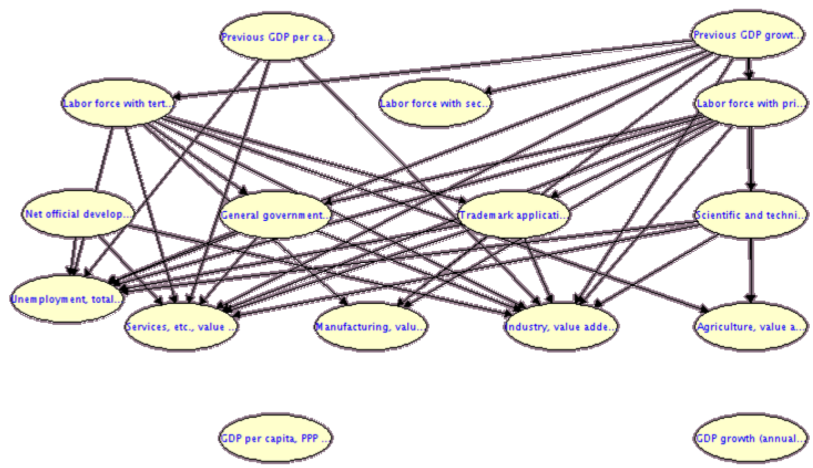
Antigua and Barbuda



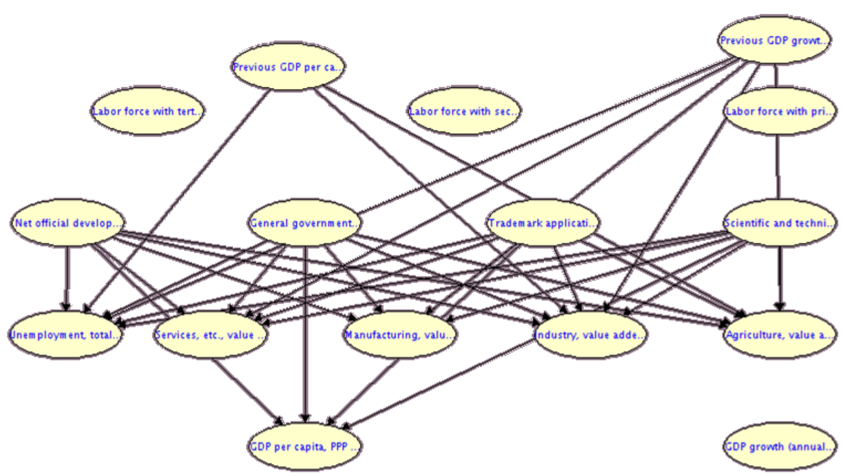
Australia



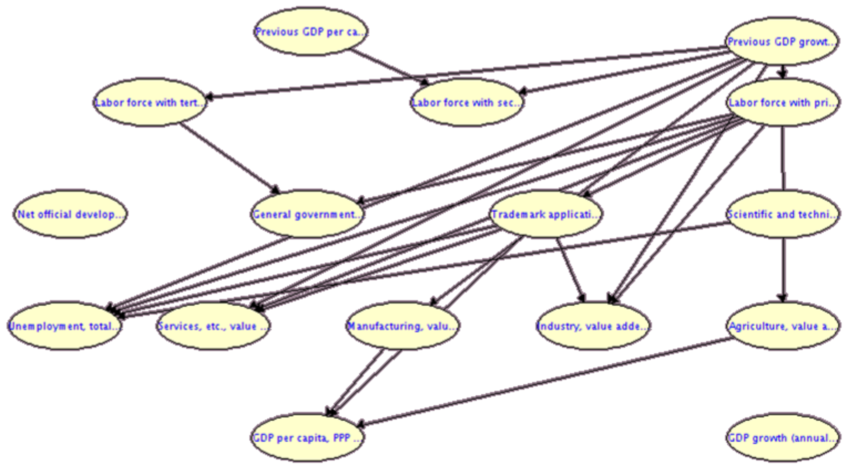
Austria



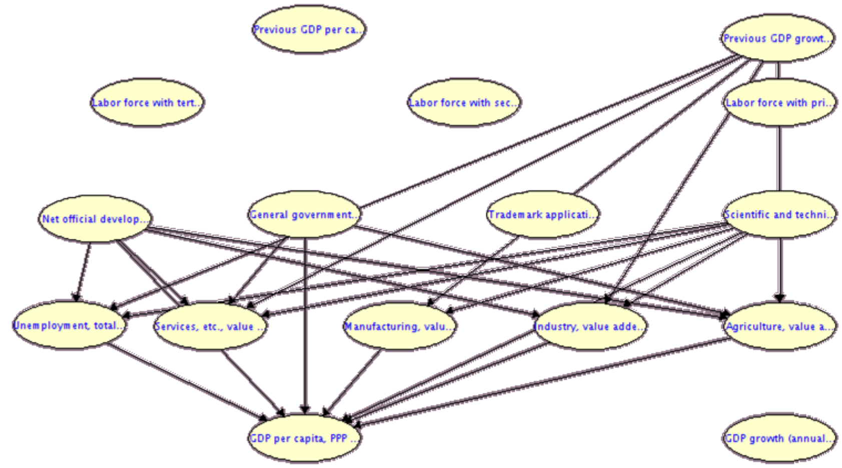
Azerbaijan



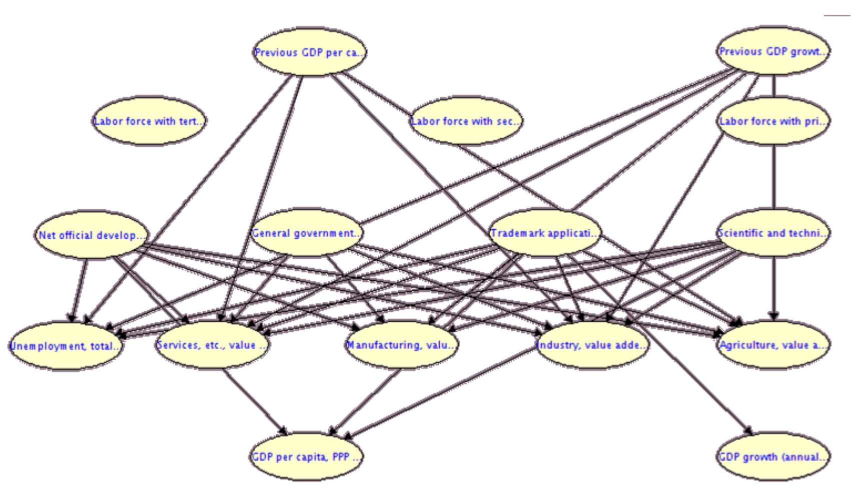
Burundi



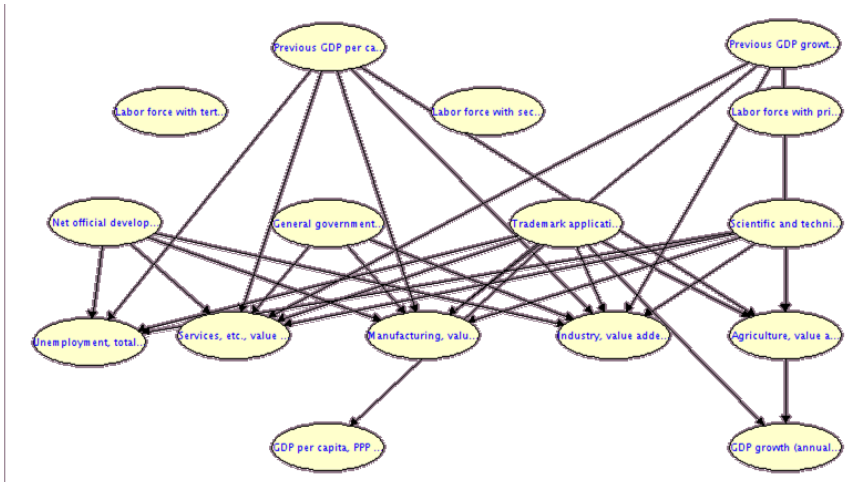
Belgium



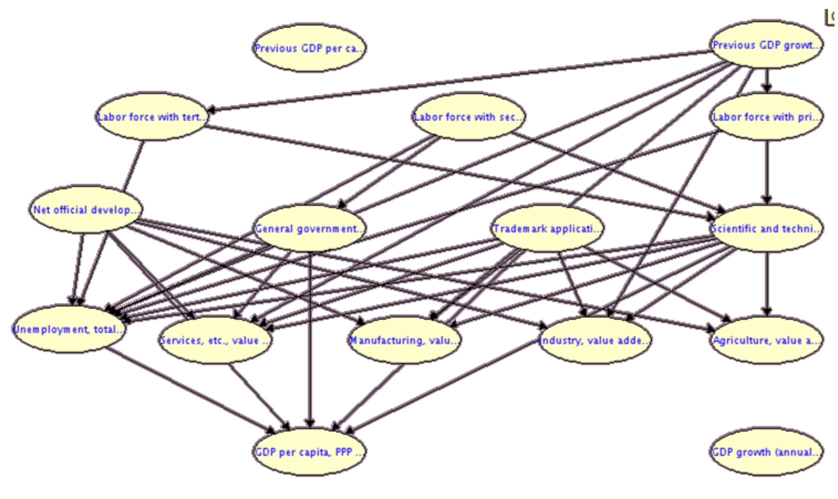
Benin



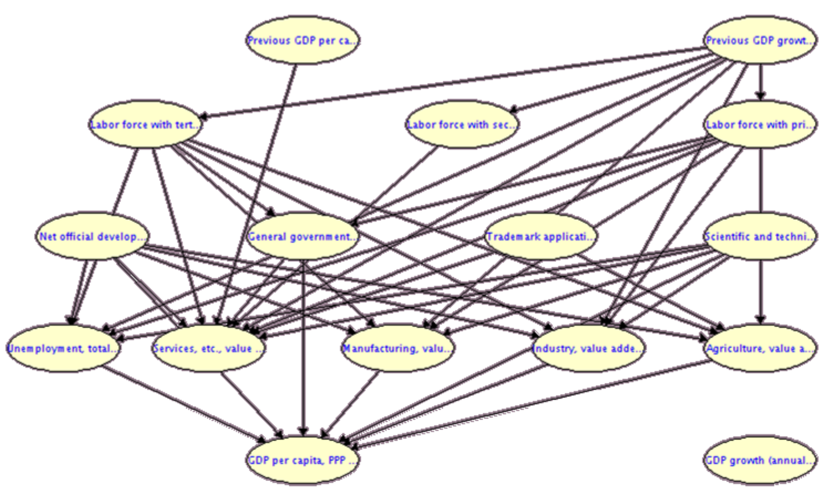
Burkina Faso



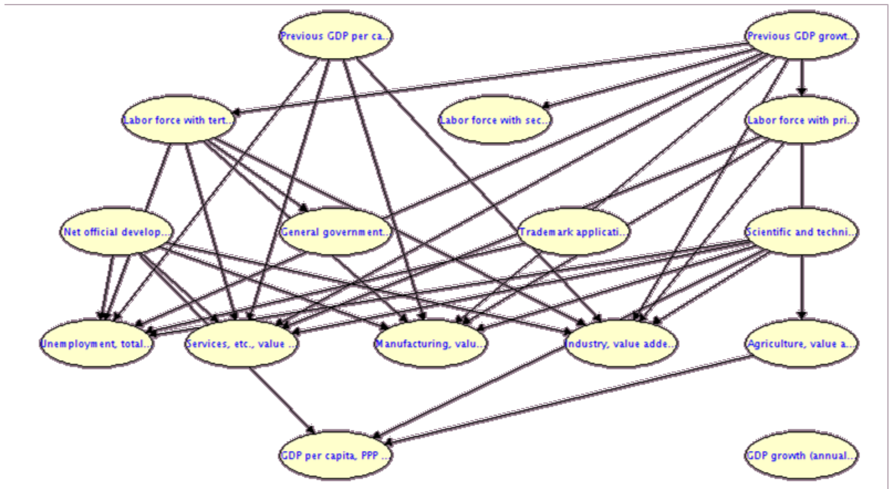
Bangladesh



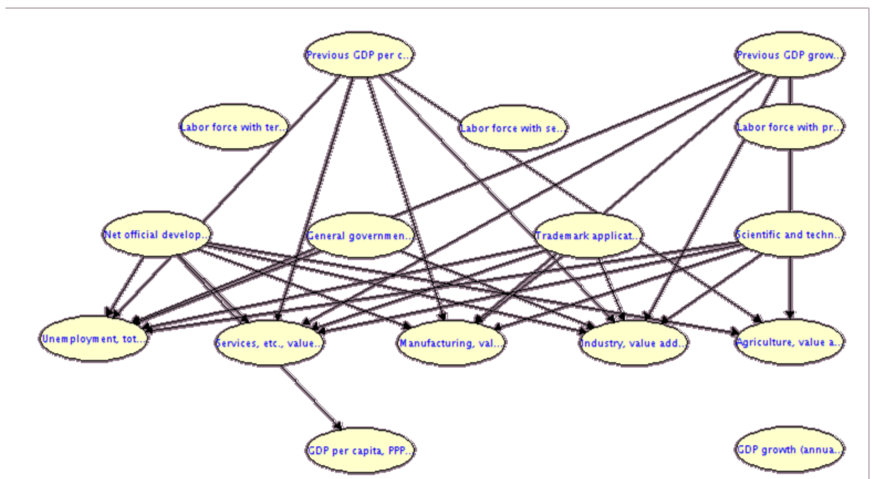
Bahrain



Bahamas, The

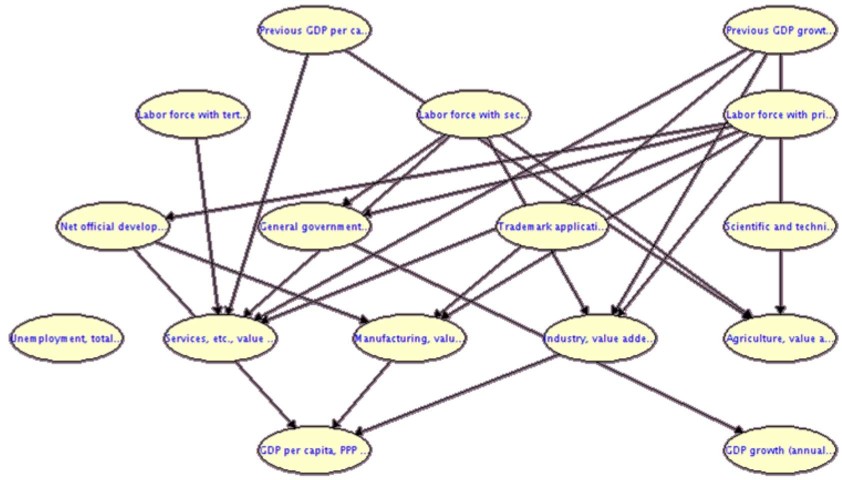


Bosnia and Herzegovina

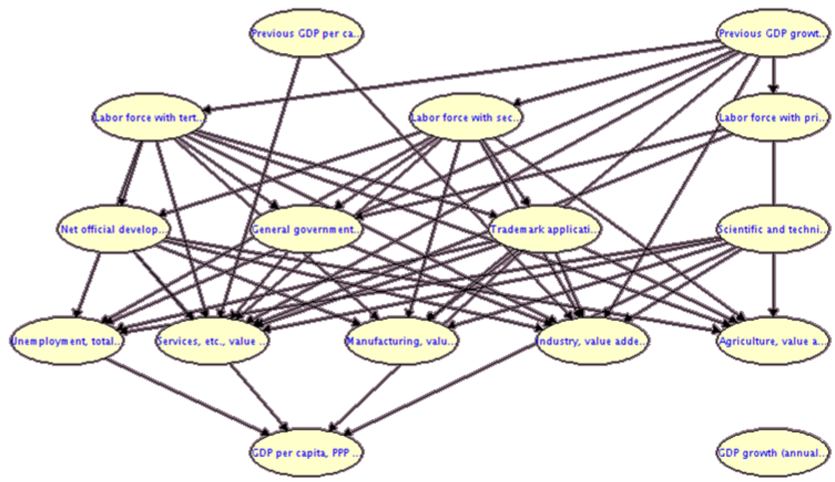


Belarus

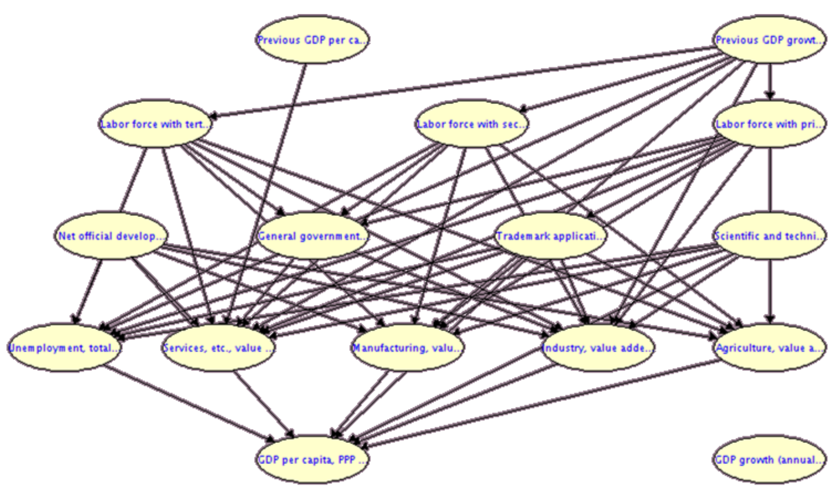
Belize



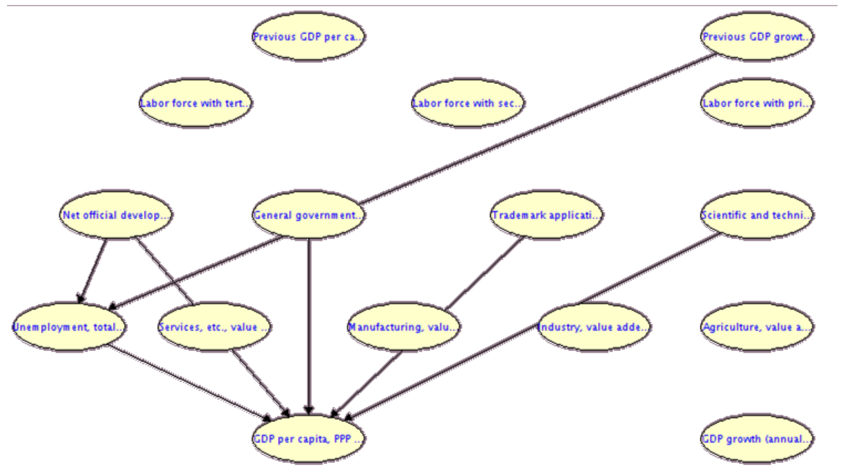
Bermuda



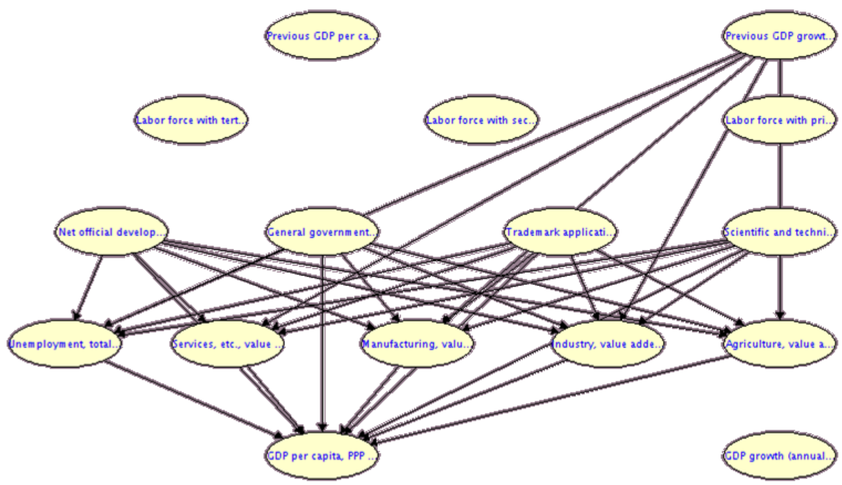
Bolivia



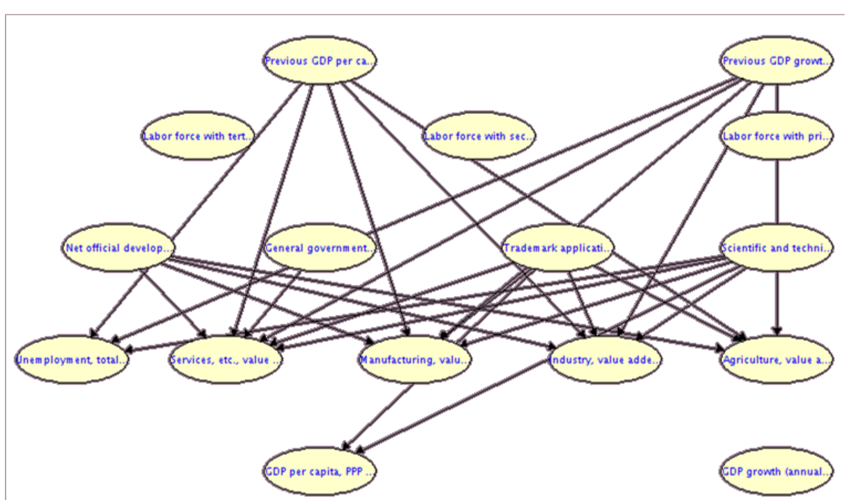
Brazil



Barbados

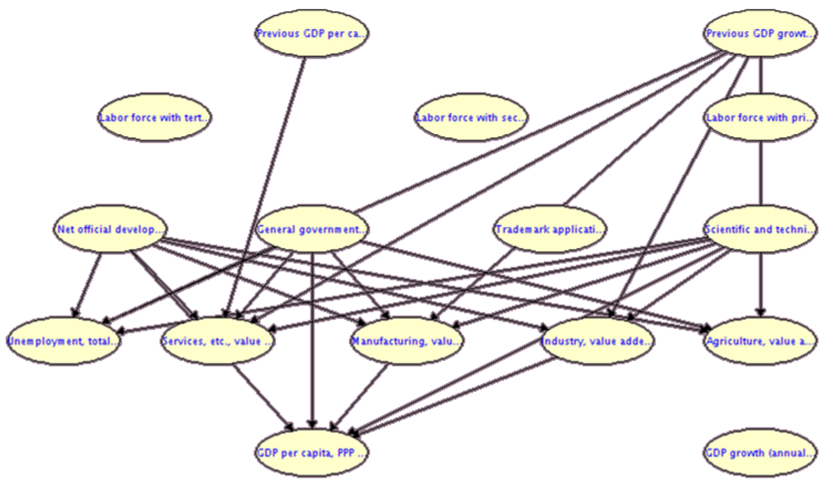


Brunei Darussalam

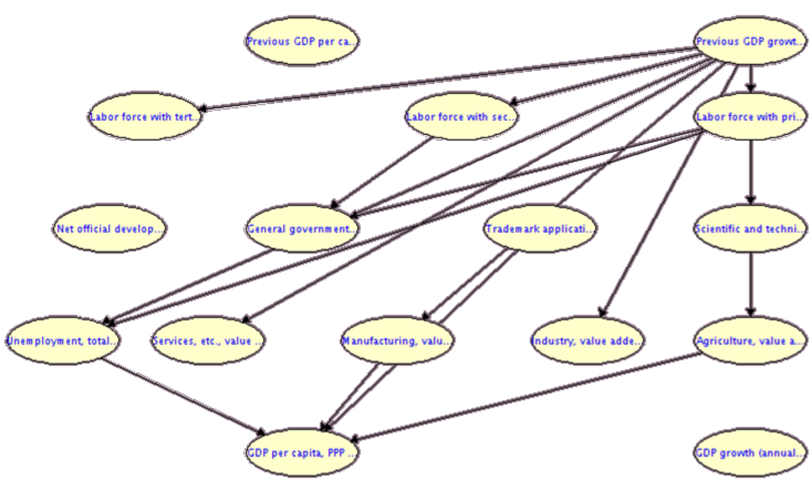


Bhutan

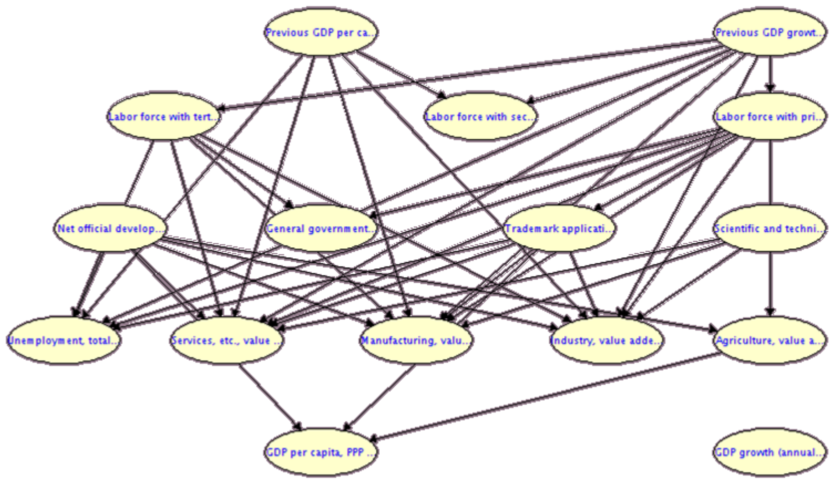
Botswana



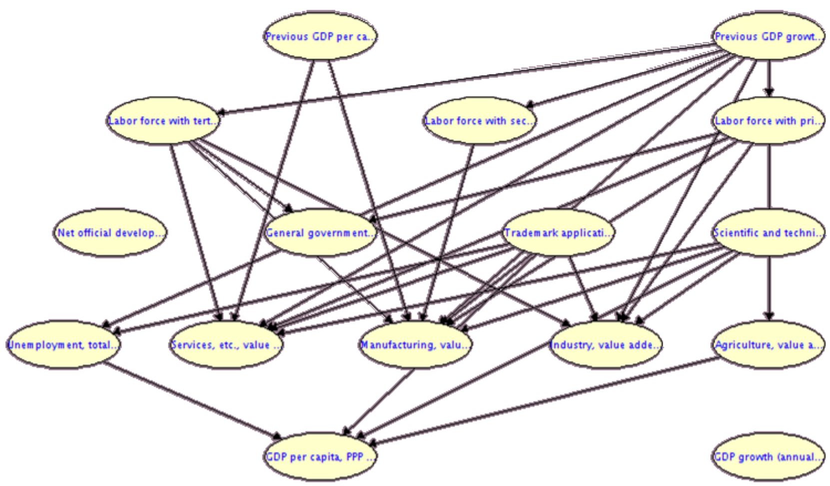
Central African Republic



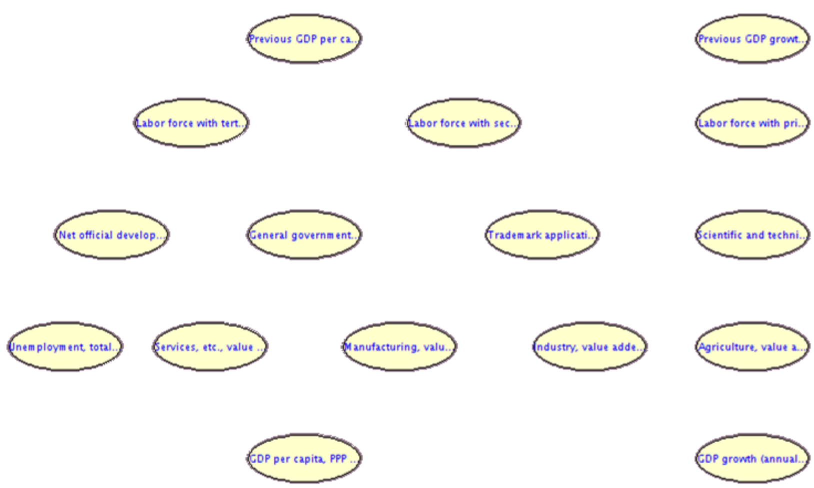
Canada



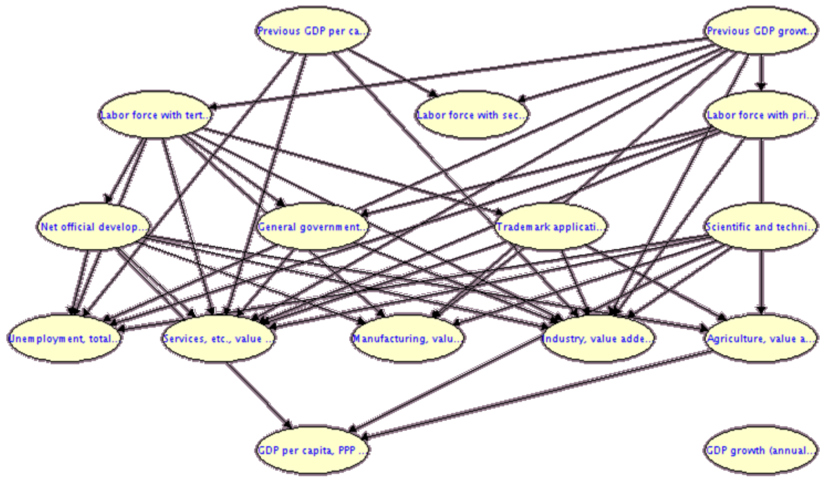
Central Europe and the Baltics



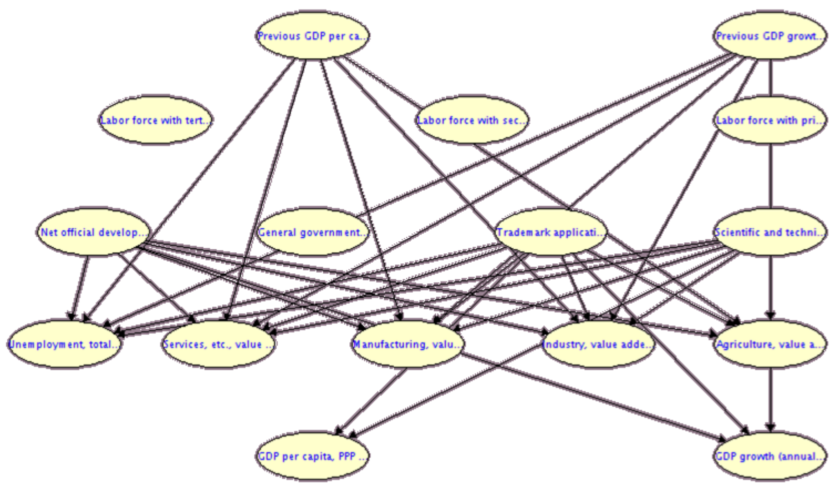
Switzerland



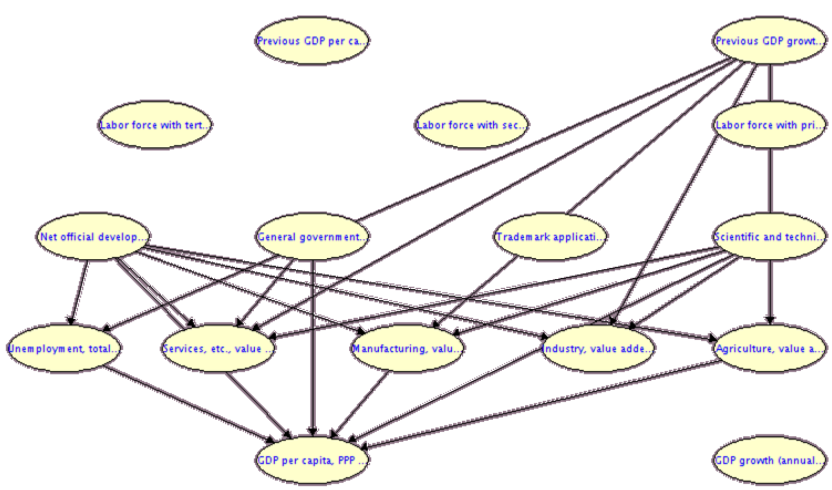
Channel Islands



Chile

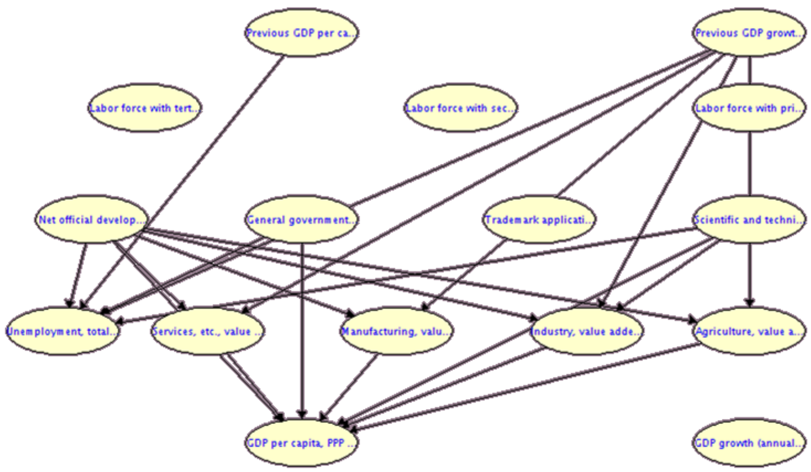


China

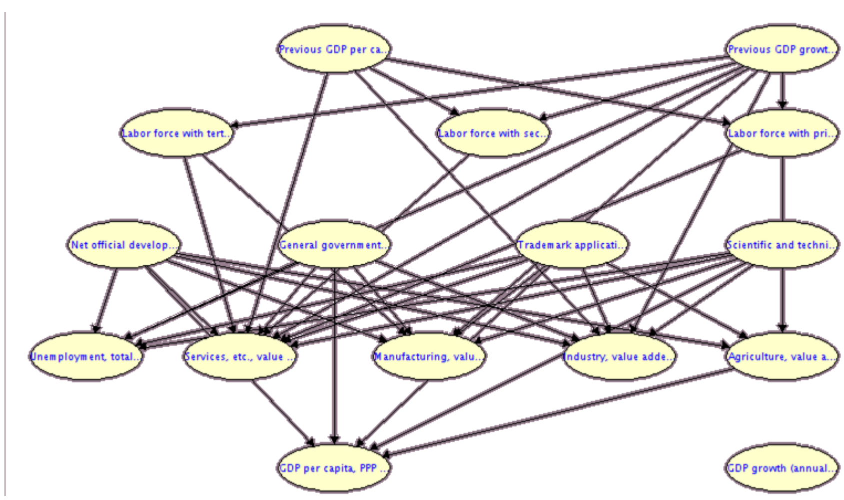


Côte d'Ivoire

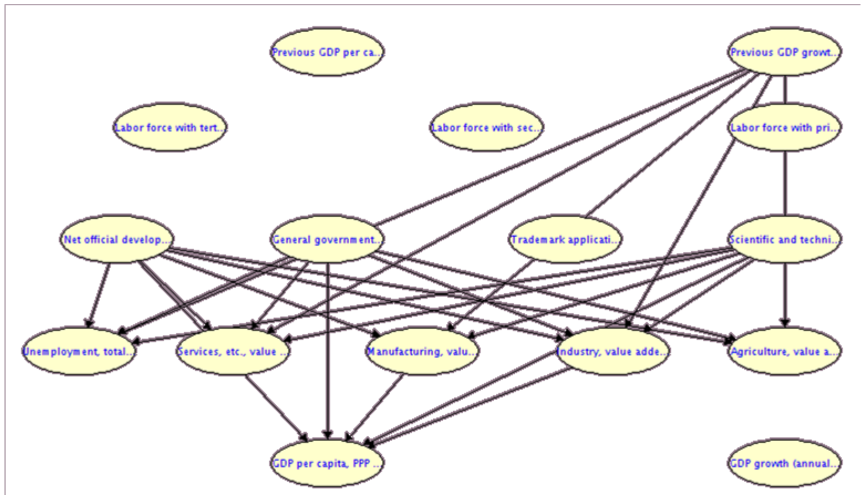
Cameroon



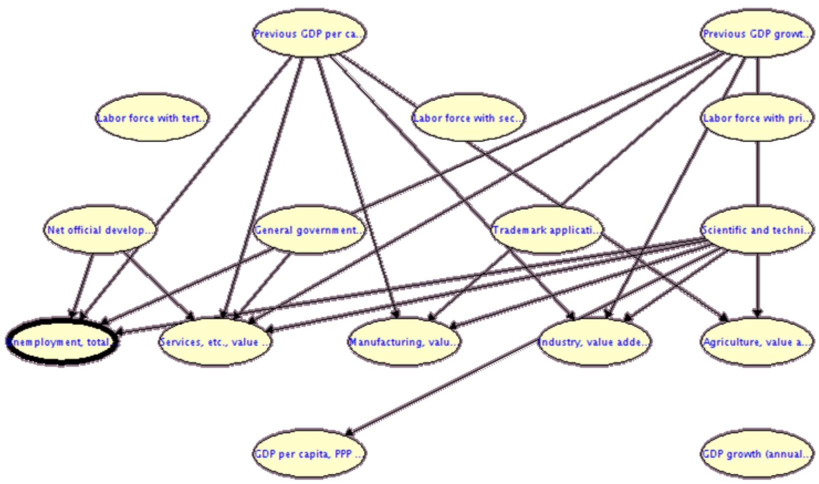
Congo, Rep.



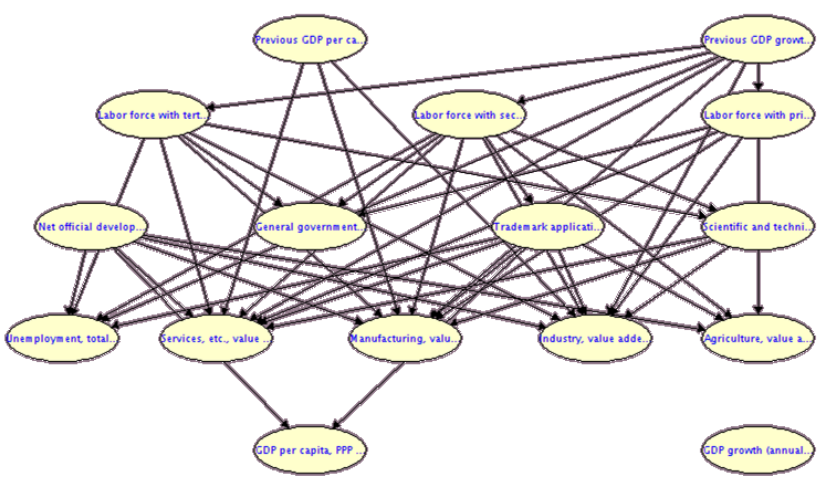
Colombia



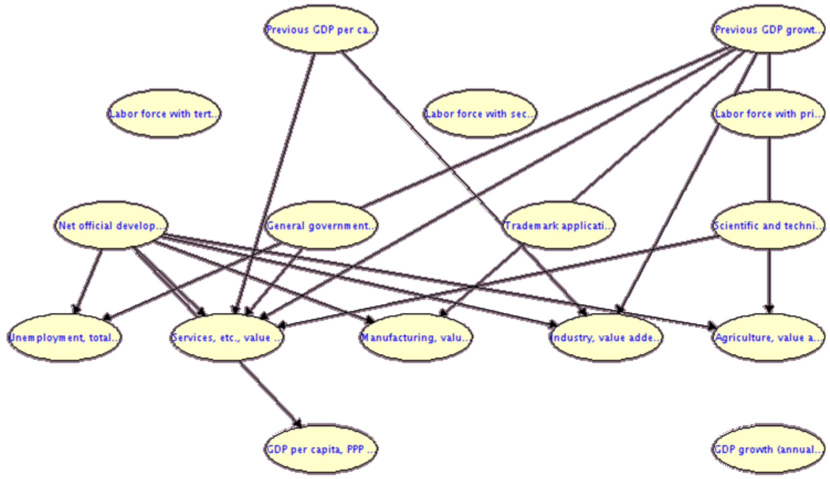
Comoros



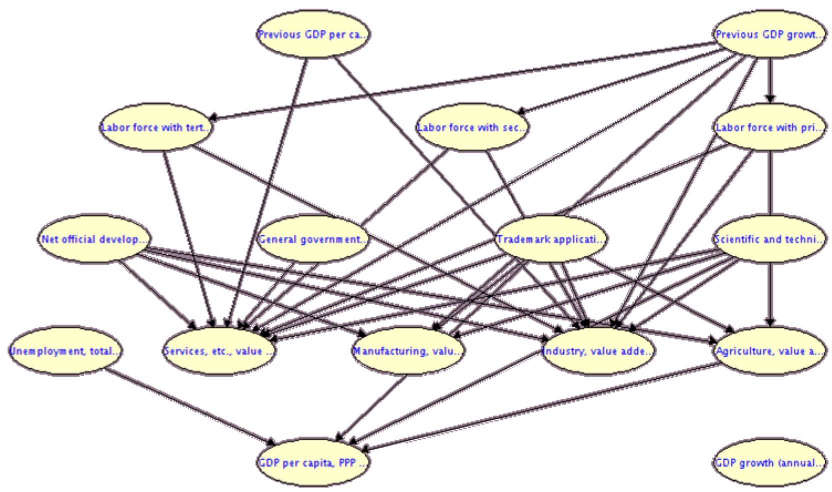
Cabo Verde



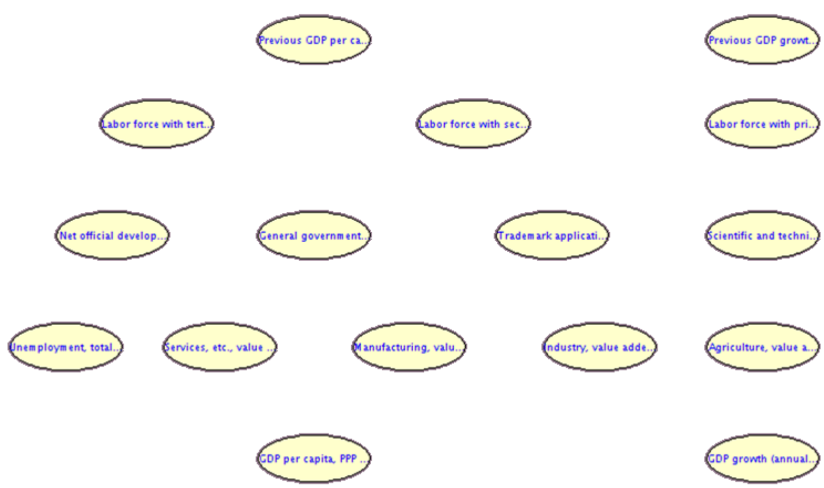
Costa Rica



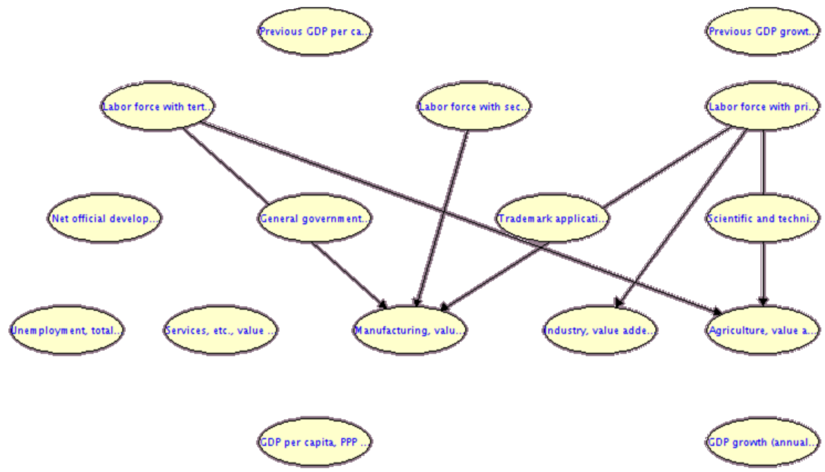
Caribbean small states



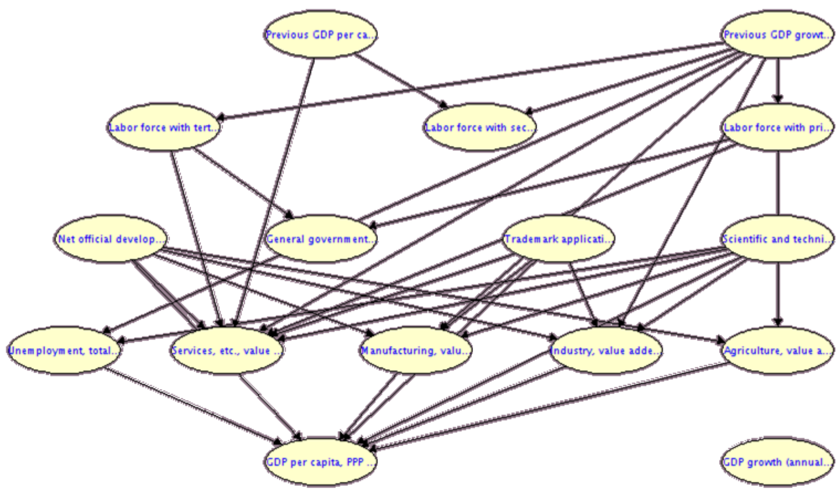
Cuba



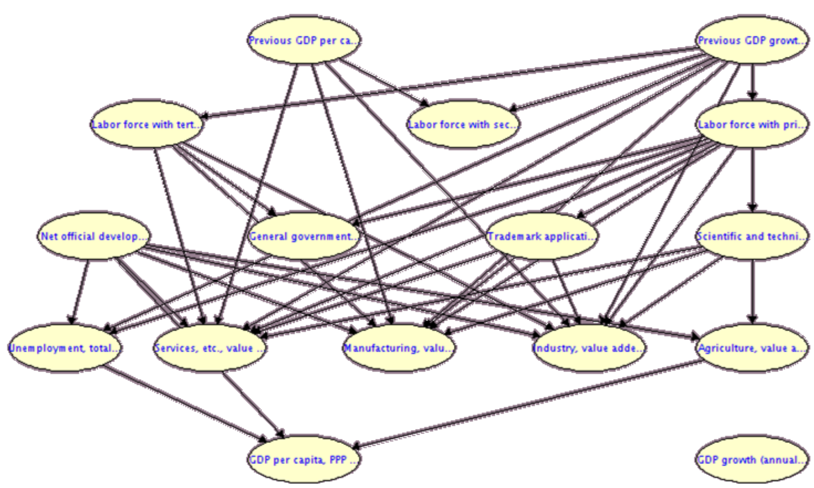
Curaçao



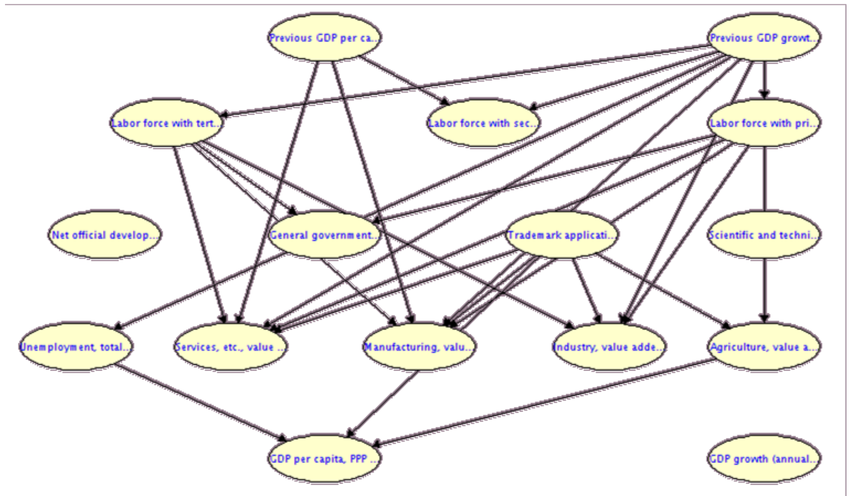
Cayman Islands



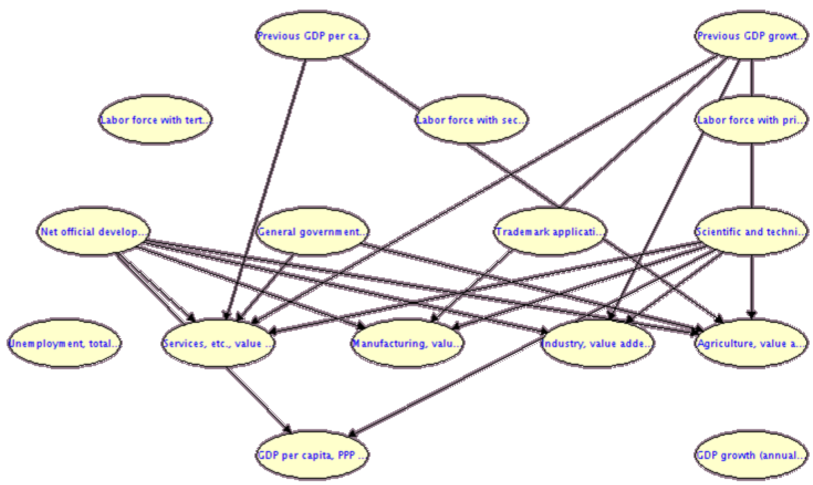
Cyprus



Czech Republic

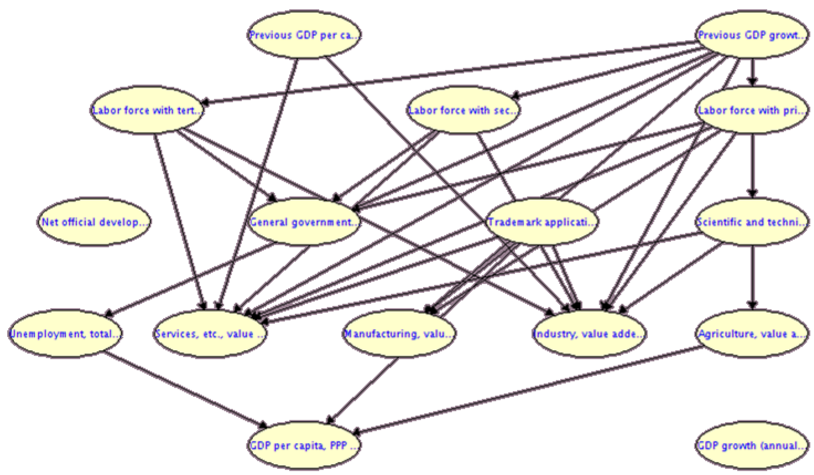


Germany



Djibouti

Dominica



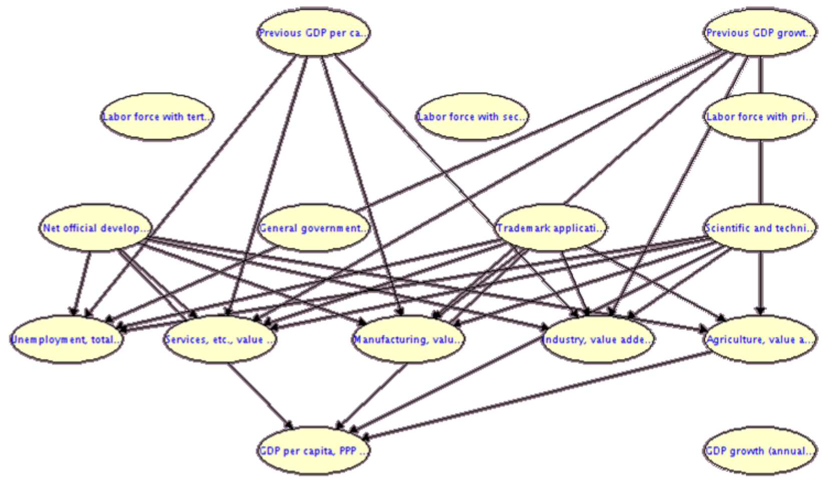
Denmark

Dominican Republic

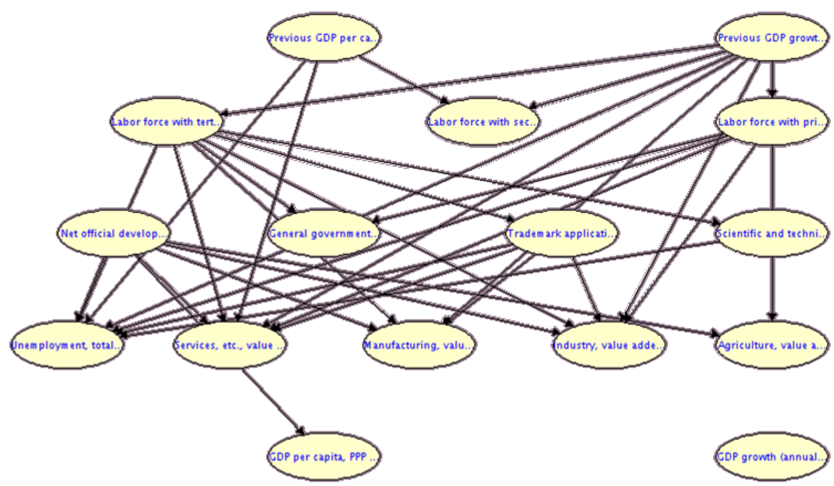
Algeria

East Asia & Pacific (developing only)

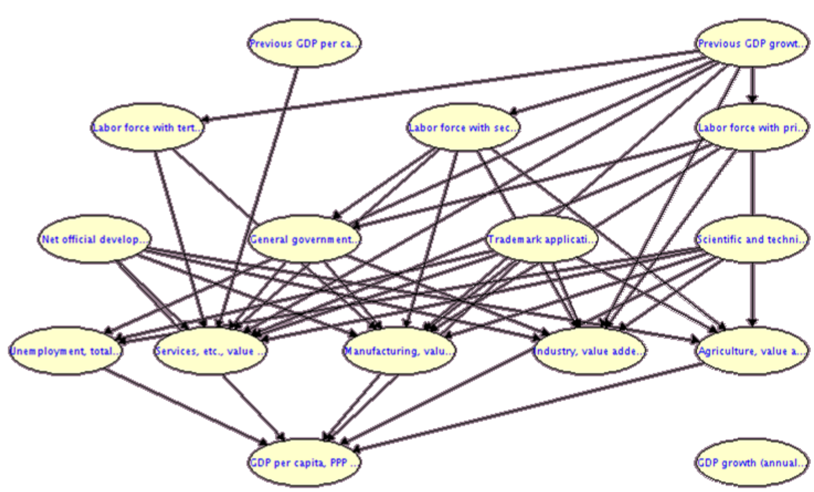
East Asia & Pacific (all income levels)



Europe & Central Asia (developing only)



Europe & Central Asia (all income levels)

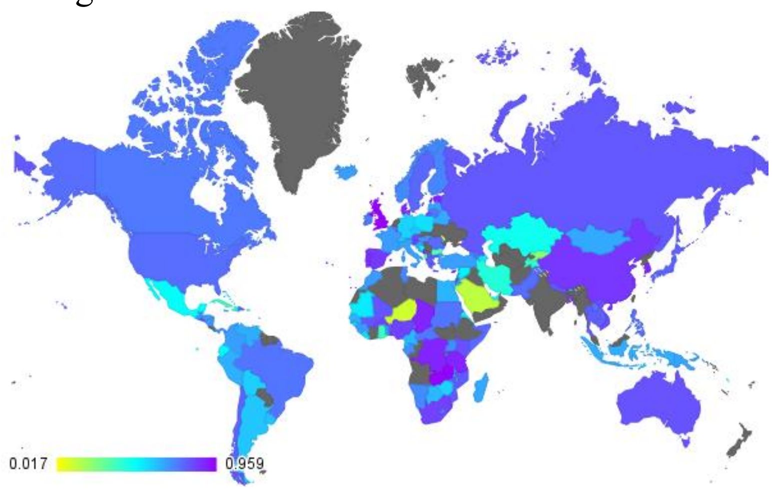


Ecuador

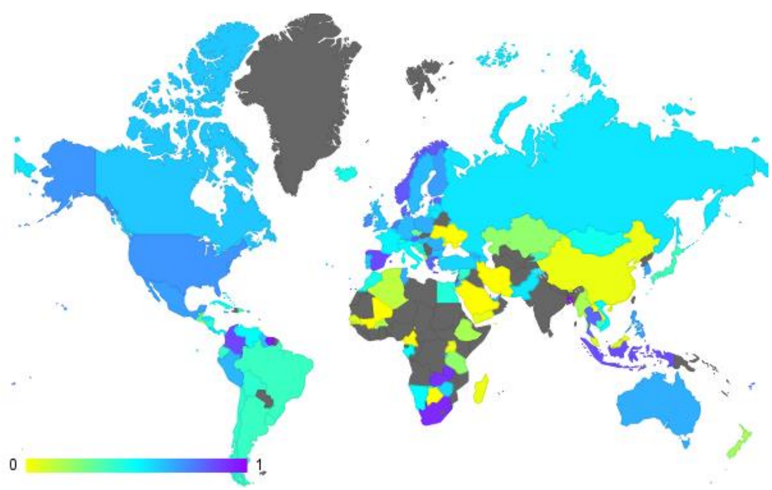
5. GEOGRAPHIC DISTRIBUTION OF ACCURACY RESULTS FOR EACH VARIABLE

This appendix displays the average accuracy result obtained—when estimating each of the variables—for the network of each country. Countries are shown in gray when there were not enough values for that variable for any of the random splits to contain a "correct" value for it.

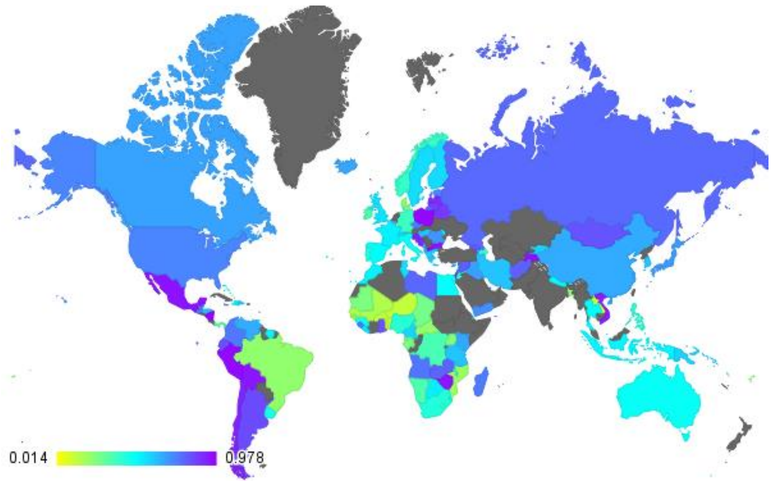
5.1. Accuracy results for the models generated with the Smets and Wouters Domain Knowledge model



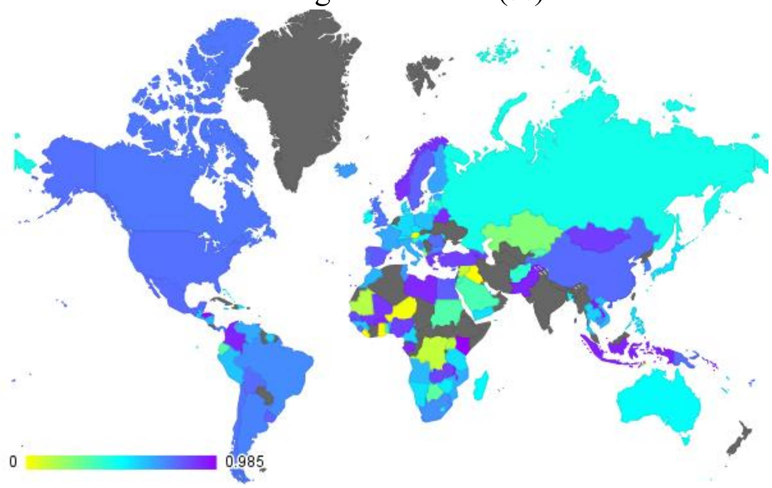
Final consumption expenditure (constant LCU)



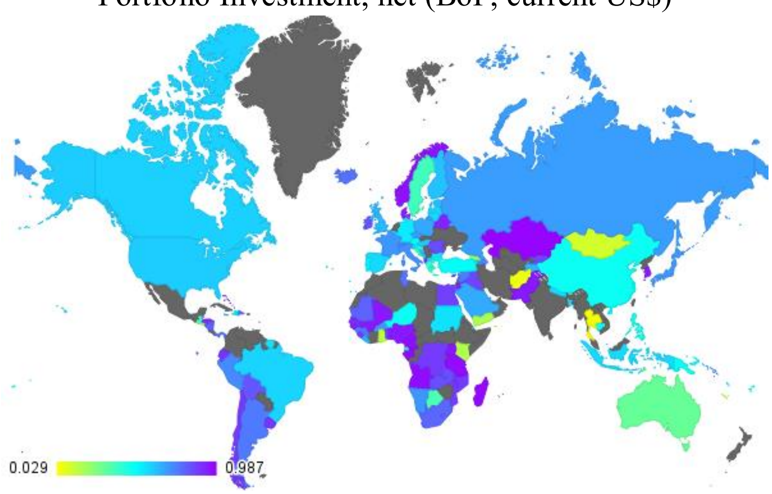
Wage and salaried workers, total (% of total employed)



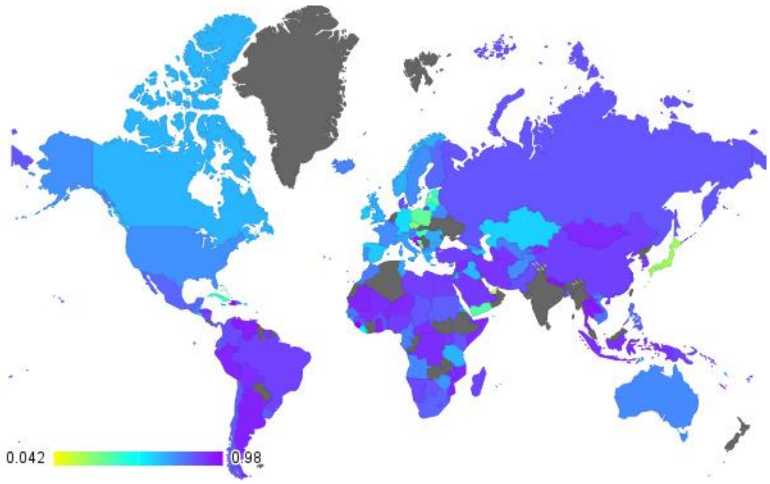
Lending interest rate (%)



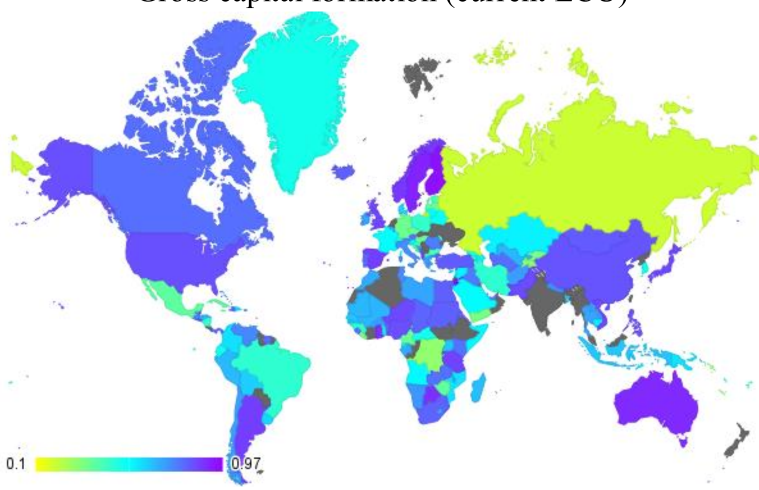
Portfolio Investment, net (BoP, current US$)



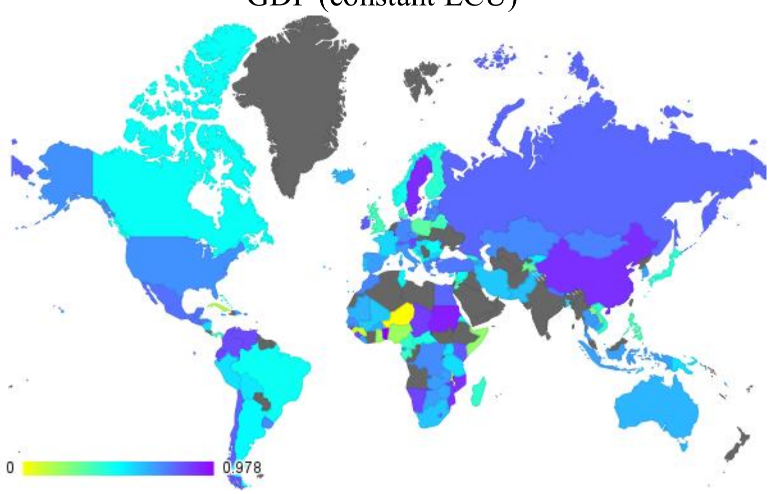
Net capital account (BoP, current US$)



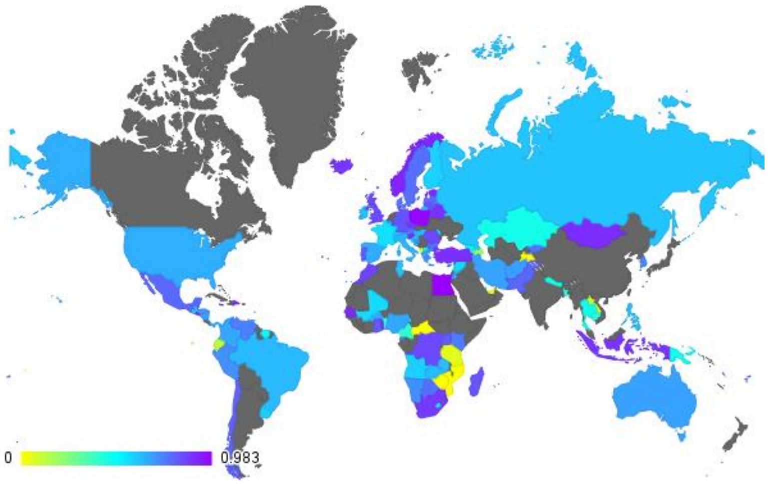
Gross capital formation (current LCU)



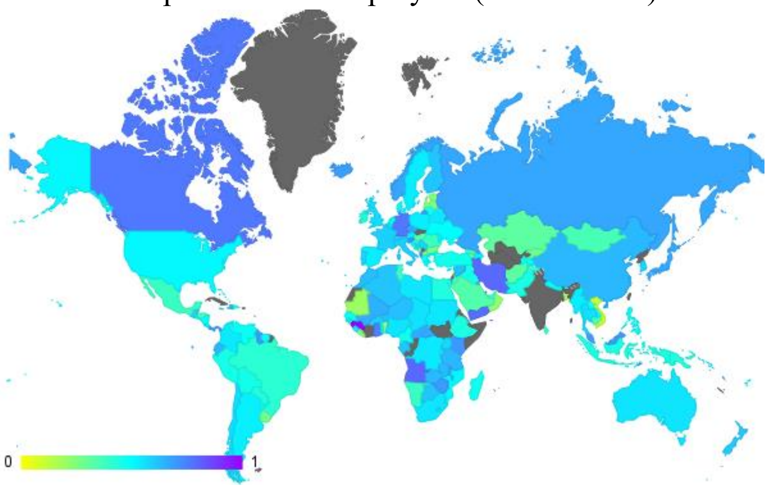
GDP (constant LCU)



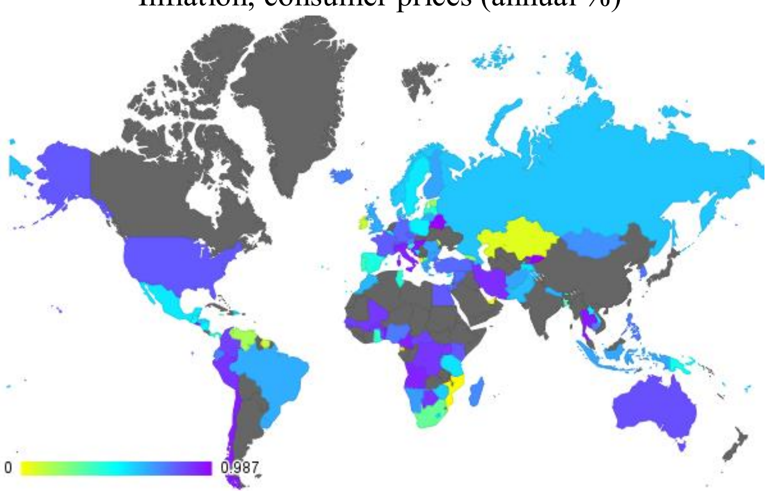
Exogenous spending



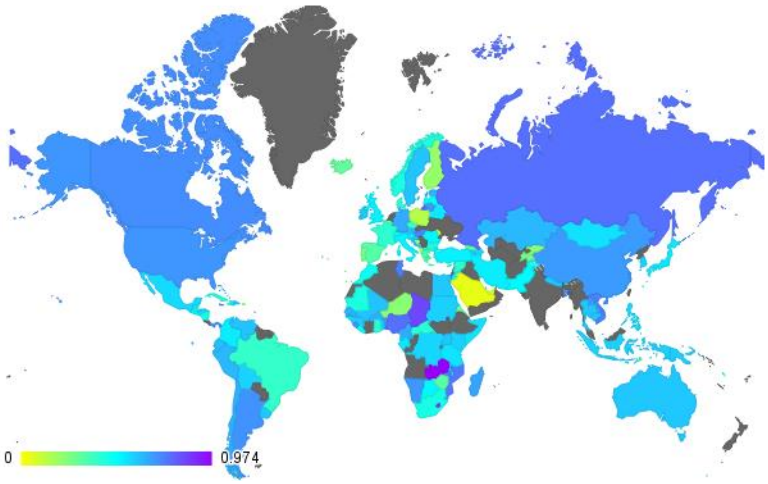
Compensation of employees (current LCU)



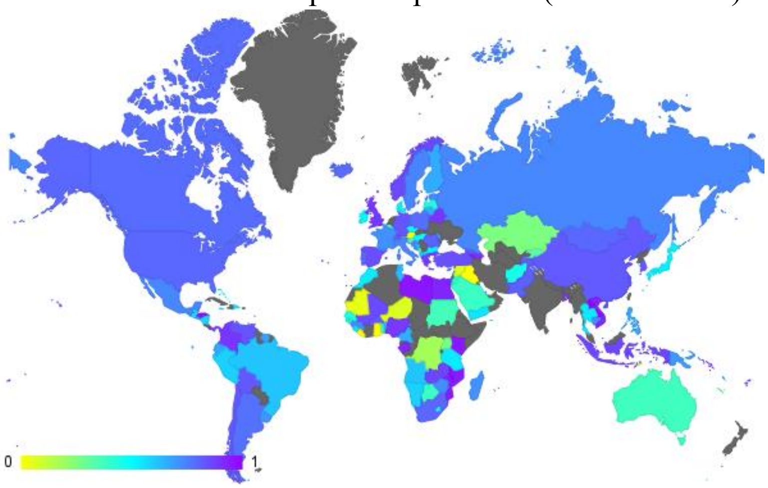
Inflation, consumer prices (annual %)



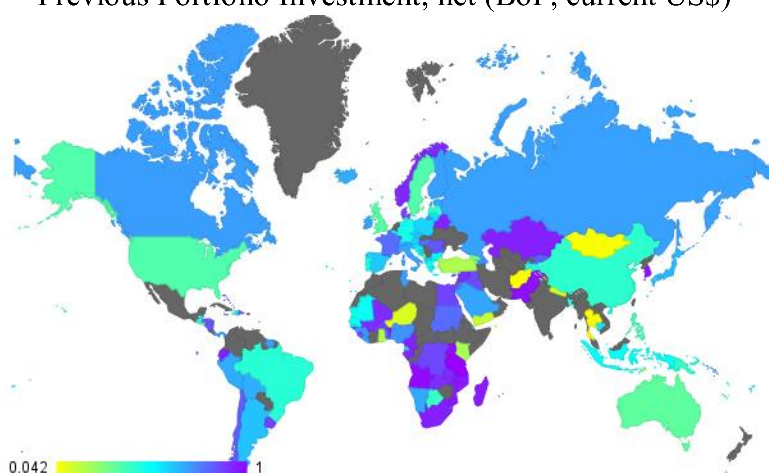
Capital-labor ratio



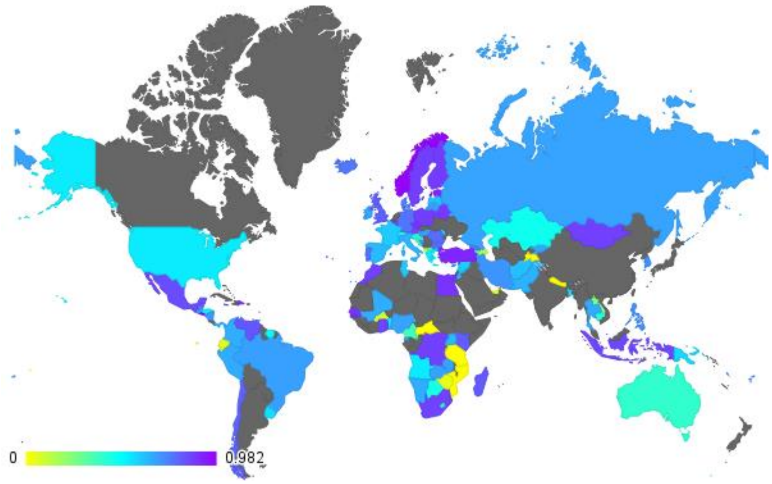
Previous Final consumption expenditure (constant LCU)



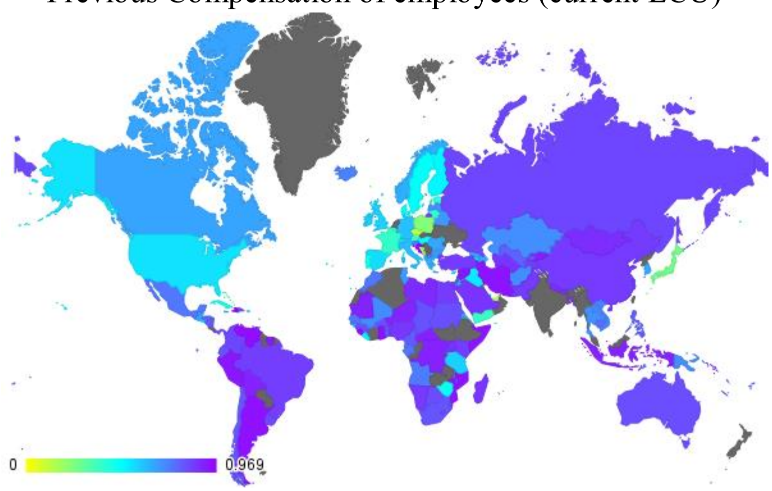
Previous Portfolio Investment, net (BoP, current US$)



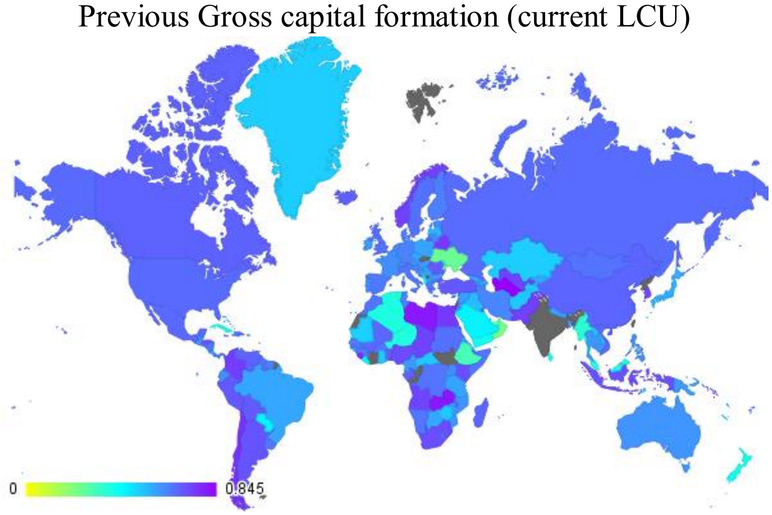
Previous Net capital account (BoP, current US$)



Previous Compensation of employees (current LCU)

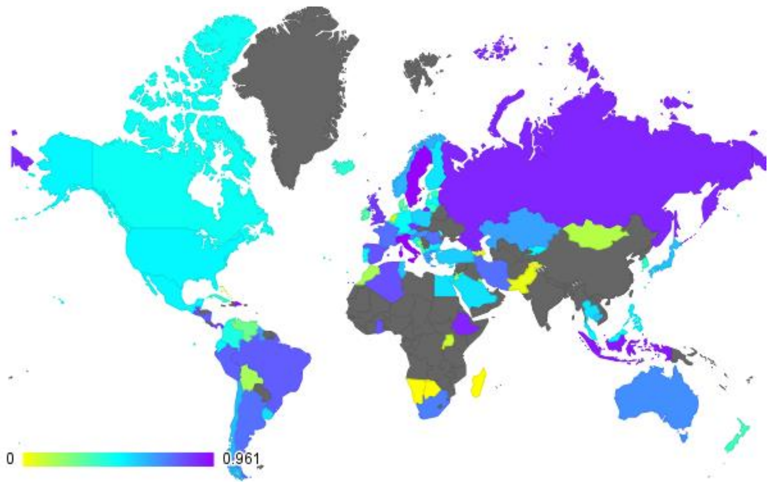


Previous Gross capital formation (current LCU)

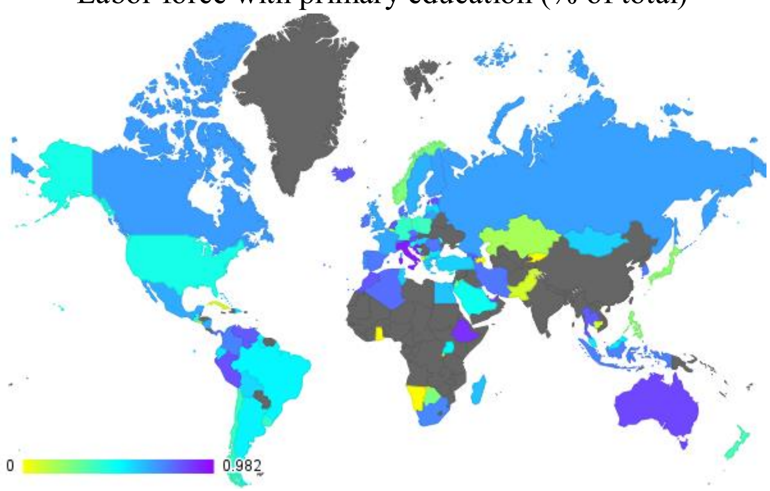


Average

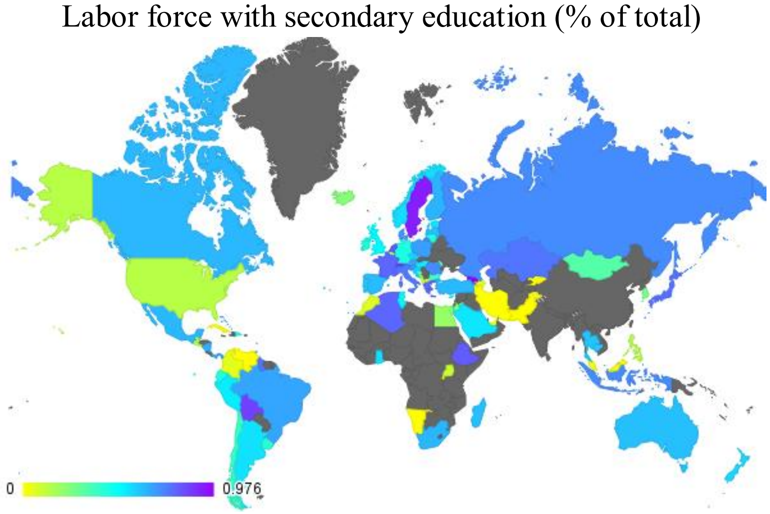
5.2. Accuracy results for the models generated with the UNESCO Domain Knowledge model



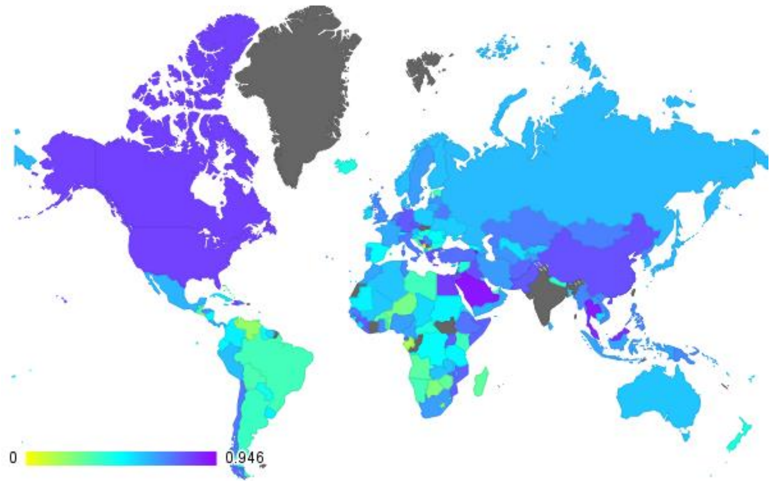
Labor force with primary education (% of total)



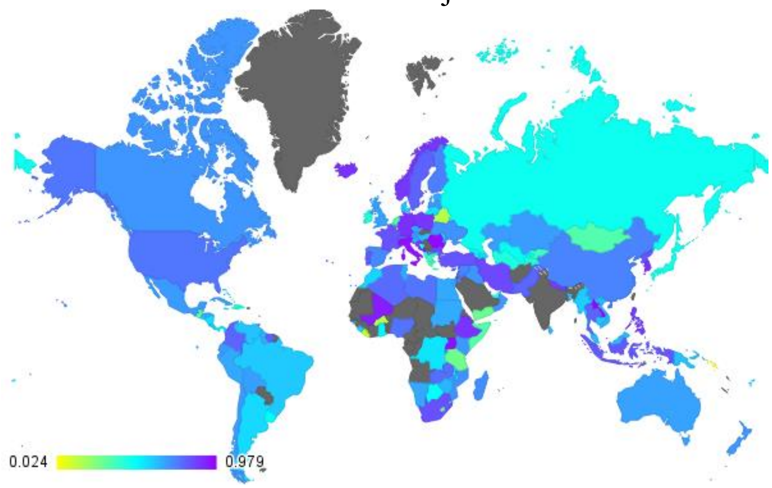
Labor force with secondary education (% of total)



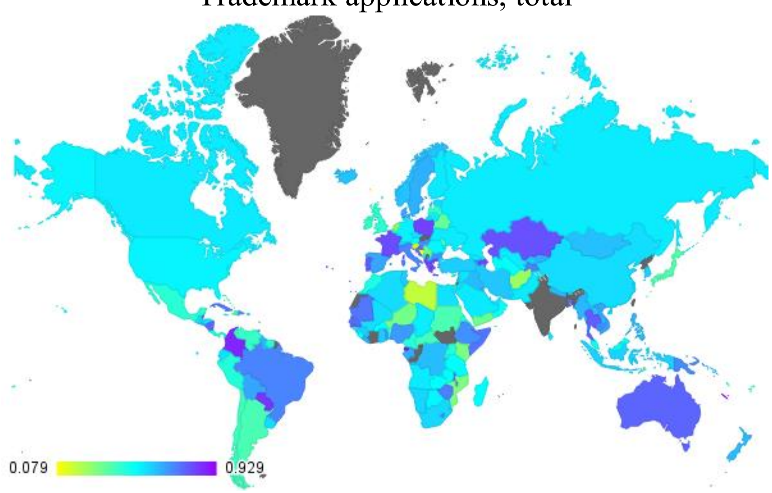
Labor force with tertiary education (% of total)



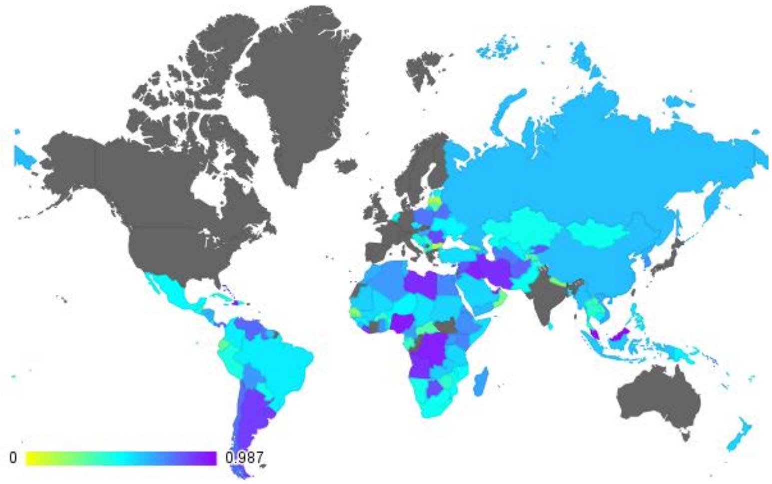
Scientific and technical journal articles



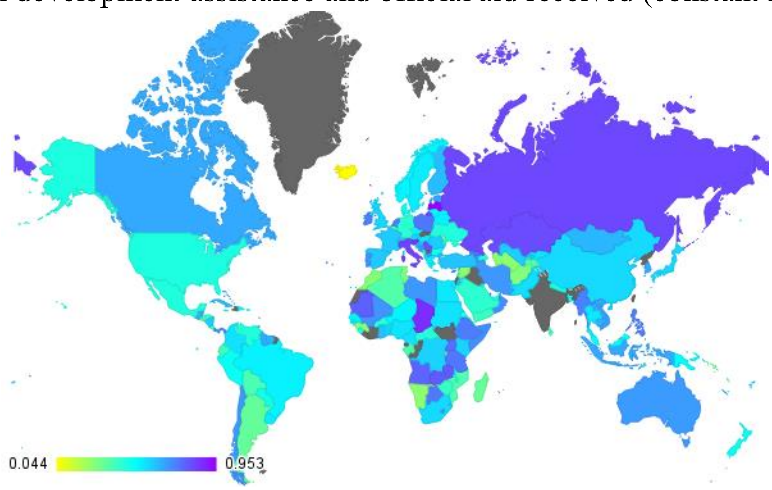
Trademark applications, total



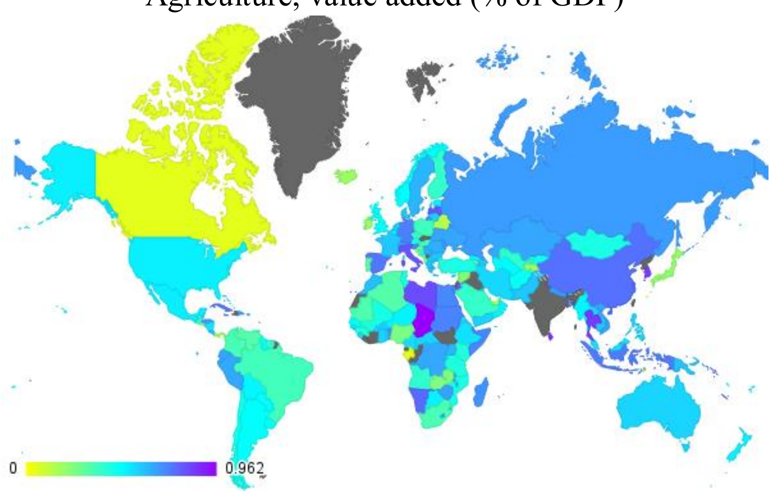
General government final consumption expenditure (% of GDP)



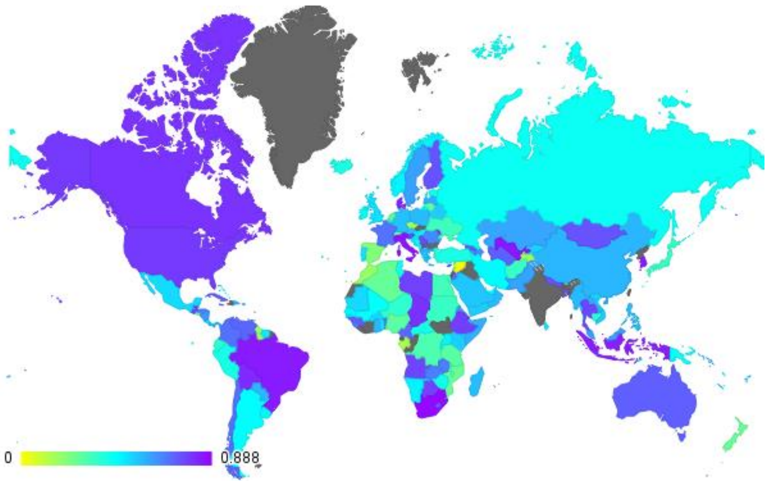
Net official development assistance and official aid received (constant 2012 US$)



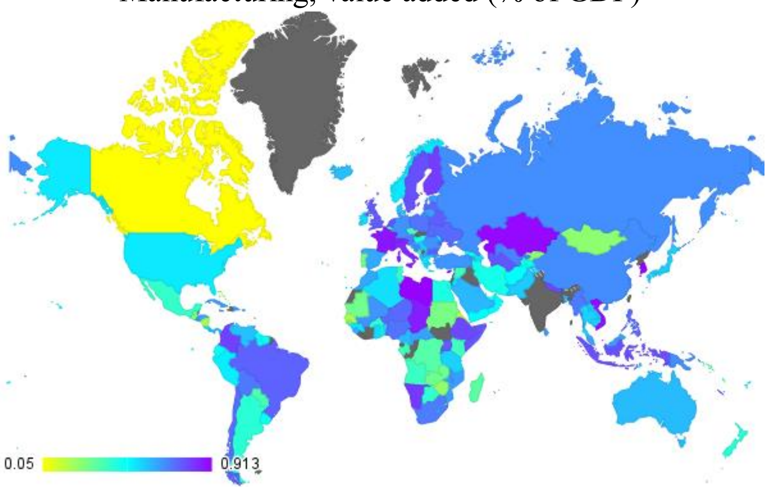
Agriculture, value added (% of GDP)



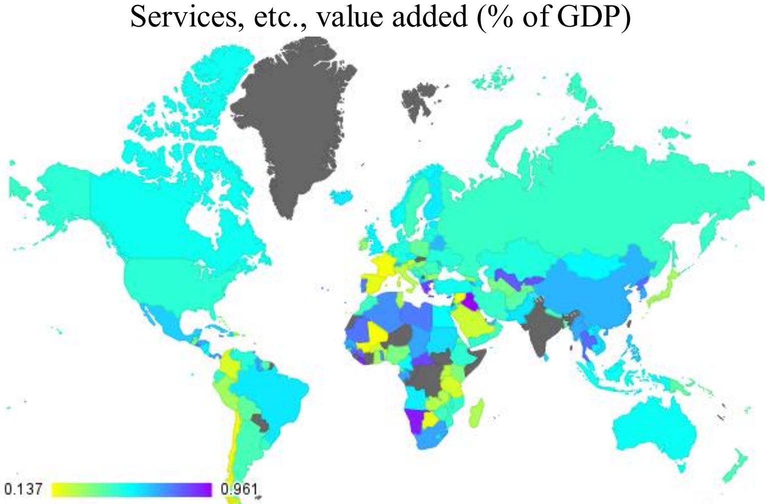
Industry, value added (% of GDP)



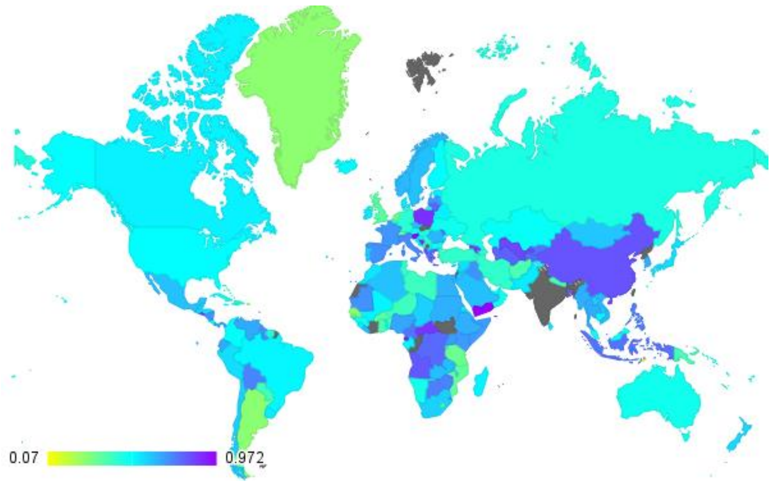
Manufacturing, value added (% of GDP)



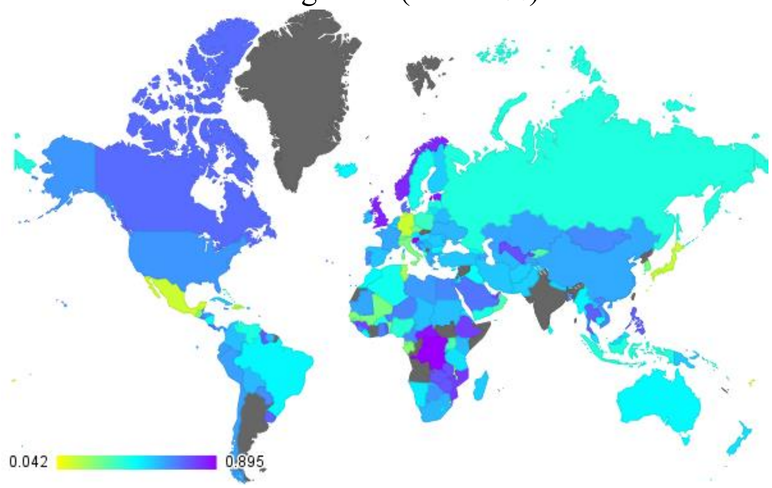
Services, etc., value added (% of GDP)



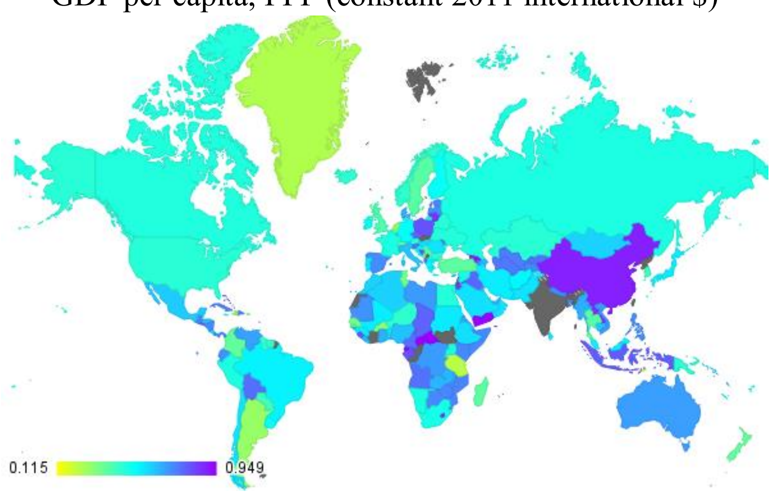
Unemployment, total (% of total labor force)



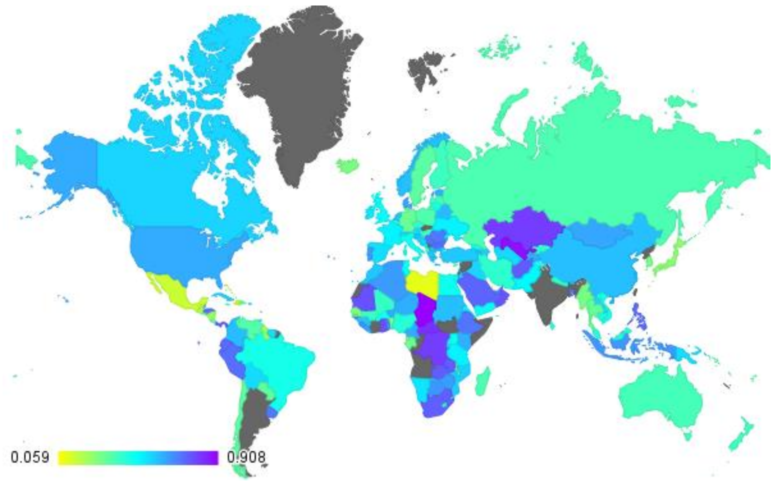
GDP growth (annual %)



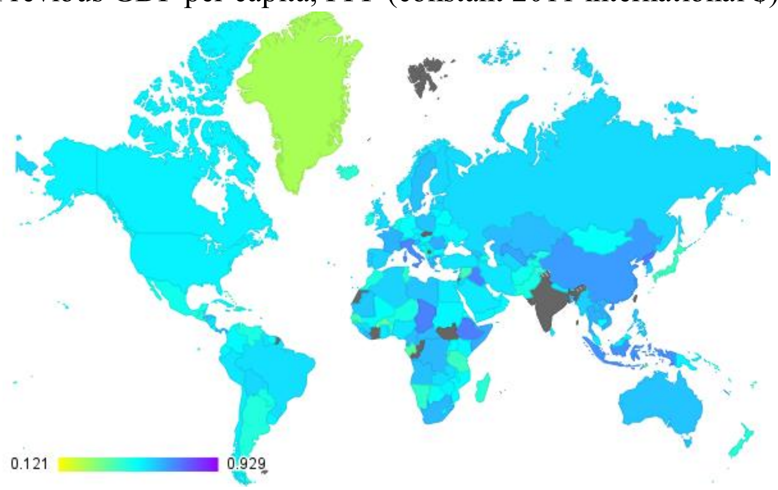
GDP per capita, PPP (constant 2011 international $)



Previous GDP growth (annual %)



Previous GDP per capita, PPP (constant 2011 international $)



Average

6. ACCURACY RESULTS FOR DIFFERENT COUNTRIES

This appendix displays the accuracy result obtained for each variable in each country. Accuracies are rounded to the nearest hundredth.

6.1. Accuracy results for the models generated with the Smets and Wouters Domain Knowledge model

Country or Group	Wage and salaried workers, total (% of total employed)	Lending interest rate (%)	Portfolio Investment, net (BoP, current US$)	Net capital account (BoP, current US$)	Gross capital formation (current LCU)	GDP (constant LCU)	Exogenous spending	Compensation of employees (current LCU)	Inflation, consumer prices (annual %)	Capital-labor ratio	Previous Final consumption expenditure (constant LCU)]	Previous Portfolio Investment, net (BoP, current US$)	Previous Net capital account (BoP, current US$)	Previous Compensation of employees (current LCU)	Previous Gross capital formation (current LCU)
Afghanistan		0.80	0.41	0.06	0.72	0.70		0.68	0.32	0.67		0.46	0.04	0.61	0.70
Albania									0.60						
Algeria	0.12	0.68	0.16	0.93	0.72	0.89	0.47	0.00	0.75	0.00	0.13	0.60	0.60	0.13	0.67
American Samoa	0.46	0.94	0.87	0.32	0.90	0.48	0.65	0.79	0.51	0.98	0.67	0.85	0.39	0.89	0.88
Andorra	0.05								0.52						
Angola	0.13								0.47						
Antigua and Barbuda	0.54								0.15						
Arab World	0.13								0.62						
Argentina	0.00														
Armenia						0.53									
Aruba		0.75	0.63	0.96	0.73	0.54		0.58	0.80	0.92		0.61	0.98	0.53	0.74
Australia		0.26	0.63	0.96	0.71	0.23		0.31	0.25	0.38		0.68	0.91	0.68	0.51
Austria	0.39	0.84	0.73	0.76	0.92	0.86	0.49		0.53		0.69	0.78	0.64		0.94
Azerbaijan	0.59	0.84	0.93	0.58	0.55	0.39	0.65	0.41	0.15	0.86	0.58	0.94	0.42	0.42	0.53
Bahamas, The	0.79	0.74	0.43	0.93	0.76	0.46			0.55			0.67	0.90		0.58
Bahrain	0.66	0.48	0.49	0.32	0.73	0.90	0.63	0.67	0.55	0.83	0.59	0.38	0.34	0.39	0.82
Bangladesh	0.81		0.00	0.50	0.76	0.78	0.86	0.77	0.55	0.74	0.62	0.00	0.61	0.85	0.38
Barbados	0.59	0.42	0.86	0.73	0.60	0.57	0.68	0.19	0.53	0.63	0.57	0.90	0.74	0.17	0.66
Belarus	0.54	0.46	0.46	0.89	0.30	0.71	0.45		0.52		0.61	0.49	0.80		0.33
Belgium	0.67	0.28	0.90	0.92	0.74	0.78	0.61	0.63	0.45	0.41	0.29	0.86	0.93	0.63	0.71
Belize	1.00	0.22	0.51	0.59	0.85	0.63	0.60	0.44	0.18	0.34	0.71	0.90	0.46	0.58	0.88
Benin	0.57	0.74	0.75	0.63	0.69	0.28		0.05	0.37	0.09	0.59	0.90	0.85	0.21	0.49
Bermuda		0.87	0.90	0.92	0.73	0.52	0.36	0.88	0.55	0.98	0.53	0.89	0.93	0.88	0.68
Bhutan	0.60	0.68	0.48	0.61	0.84	0.69	0.58	0.71	0.58	0.47	0.30	0.33	0.50	0.89	0.40
Bolivia	0.17	0.64	0.77	0.76	0.62	0.84	0.26	0.88	0.60	0.62	0.29	0.78	0.71	0.77	0.62
Bosnia and Herzegovina		0.15	0.68	0.94	0.87	0.78	0.97	0.39	0.19	0.95	0.70	0.65	0.97	0.54	0.87
Botswana	0.11		0.31		0.50	0.62						0.19			0.60
Brazil	0.05	0.49		0.39	0.79	0.85	0.36	0.74	0.42	0.44	0.33		0.52	0.73	0.80
Brunei Darussalam	0.41	0.93	0.78	0.82	0.89	0.62	0.53		0.46		0.57	0.83	0.67		0.90
Bulgaria	0.55	0.89	0.72	0.65	0.31	0.72	0.63	0.74	0.60	0.56	0.20	0.66	0.64	0.71	0.21
Burkina Faso	0.00	0.53	0.31	0.37	0.83	0.89	0.67	0.77	0.62	0.88	0.50	0.38	0.42	0.45	0.82
Burundi	0.39	0.22	0.71	0.59	0.86	0.46	0.47	0.62	0.42	0.65	0.38	0.61	0.44	0.67	0.85
Cabo Verde			0.46	0.52	0.65	0.67	0.44		0.56		0.51	0.44	0.38		0.66

Country or Group	Wage and salaried workers, total (% of total employed)	Lending interest rate (%)	Portfolio Investment, net (BoP, current US$)	Net capital account (BoP, current US$)	Gross capital formation (current LCU)	GDP (constant LCU)	Exogenous spending	Compensation of employees (current LCU)	Inflation, consumer prices (annual %)	Capital-labor ratio	Previous Final consumption expenditure (constant LCU)]	Previous Portfolio Investment, net (BoP, current US$)	Previous Net capital account (BoP, current US$)	Previous Compensation of employees (current LCU)	Previous Gross capital formation (current LCU)
Cambodia	0.61	0.95	0.72	0.56	0.69	0.41	0.58	0.78	0.26	0.68	0.44	0.53	0.52	0.77	0.68
Cameroon	0.17	0.14	0.91	0.71	0.85	0.71	0.66	0.66	0.59	0.90	0.64	0.82	0.88	0.10	0.82
Canada	0.40	0.59		0.86	0.89	0.52	0.40	0.08	0.36	0.53	0.59		0.98	0.11	0.91
Caribbean small states		0.33	0.73	0.53	0.04	0.62		0.06	0.31	0.07	0.04	0.75	0.62	0.13	0.00
Cayman Islands	0.38		0.72	0.49	0.77	0.56	0.66	0.43	0.38	0.75	0.60	0.73	0.57	0.42	0.76
Central African Republic	0.00	0.58	0.57	0.96	0.76	0.35	0.40	0.41	0.47	0.92	0.65	0.65	0.93	0.37	0.79
Central Europe and the Baltics	0.20								0.47						
Chad	0.61	0.67	0.75	0.59	0.65	0.78	0.48		0.76		0.70	0.79	0.70		0.66
Channel Islands	0.06					0.40									
Chile		0.22			0.89	0.50	0.44	0.00	0.67	0.86	0.41			0.00	0.91
China		0.28			0.82	0.80	0.86		0.56		0.84				0.82
Colombia						0.36									
Comoros	0.41	0.91	0.75	0.88	0.81	0.69	0.79	0.79	0.56	0.93	0.63	0.83	0.82	0.79	0.80
Congo, Dem. Rep.	0.04	0.65	0.77	0.49	0.86	0.83	0.89		0.64		0.68	0.82	0.44		0.86
Congo, Rep.	0.86	0.75	0.91		0.84	0.59	0.83	0.66	0.54	0.86	0.52	0.89		0.65	0.84
Costa Rica		0.62	0.57	0.11	0.79	0.77	0.37		0.19		0.52	0.70	0.11		0.84
Cote d'Ivoire	0.27	0.60	0.95	0.68	0.87	0.86	0.50	0.80	0.68	0.84	0.47	0.94	0.55	0.81	0.87
Croatia	0.56								0.55						
Cuba		0.16	0.71	0.88	0.88	0.59	0.14	0.63	0.55	0.89	0.54	0.62	0.84	0.62	0.82
Curacao	0.41	0.90	0.66	0.41	0.94	0.40	0.46	0.61	0.45	0.93	0.60	0.72	0.48	0.38	0.97
Cyprus	0.50				0.42	0.38	0.14				0.43				0.46
Czech Republic															
Denmark	0.35	0.66	0.74	0.86	0.68	0.60	0.49	0.74	0.50	0.48	0.36	0.73	0.82	0.69	0.55
Djibouti	0.30	0.74	0.55	0.70	0.26	0.36	0.70	0.89	0.68	0.81	0.39	0.51	0.64	0.88	0.11
Dominica		0.36	0.11	0.88	0.90	0.29	0.71	0.84	0.54	0.88	0.56	0.18	0.87	0.86	0.91
Dominican Republic	0.79	0.24	0.57	0.93	0.85	0.60	0.38	0.76	0.50	0.63	0.41	0.52	0.88	0.74	0.48
East Asia & Pacific (all income levels)		0.48		0.49	0.81	0.73	0.06		0.50				0.60		0.86
East Asia & Pacific (developing only)	0.00	0.20	0.30	0.84	0.36	0.51		0.19	0.62	0.28		0.55	0.81	0.22	0.30
Ecuador	0.39	0.60	0.53	0.79	0.77	0.59	0.65	0.93	0.40	0.59	0.59	0.55	0.81	0.89	0.82
Egypt, Arab Rep.	0.40	0.83	0.33	0.89	0.82	0.51	0.84	0.12	0.64	0.74	0.61	0.67	0.94	0.06	0.81
El Salvador	0.29	0.01	0.56	0.52					0.63			0.43	0.43		
Equatorial Guinea	0.43	0.49	0.80	0.86	0.87	0.80	0.78	0.98	0.44	0.81	0.59	0.93	0.87	0.88	0.86
Eritrea	0.18		0.69	0.18	0.61	0.84	0.75	0.88	0.48	0.92	0.49	0.69	0.32	0.98	0.56
Estonia		0.27		0.43	0.67	0.53	0.64	0.72	0.52	0.06	0.55		0.50	0.64	0.73
Ethiopia					0.71	0.81	0.50				0.29				0.67
Euro area	0.78	0.72	0.65	0.63	0.41	0.40	0.65	0.91	0.24	0.20	0.63	0.72	0.69	0.92	0.39
Europe & Central Asia (all income levels)									0.55						
Europe & Central Asia (developing only)	0.41	0.59		0.88	0.91	0.90		0.83	0.34		0.00		0.94	0.86	0.88
European Union	0.17								0.44						
Faroe Islands					0.19										0.21
Fiji	0.75	0.24	0.77	0.49	0.81	0.43		0.76	0.52	0.60		0.88	0.56	0.71	0.77
Finland	0.69	0.56	0.67	0.62	0.78	0.94	0.41	0.58	0.61	0.67	0.20	0.66	0.70	0.87	0.50

Country or Group	Wage and salaried workers, total (% of total employed)	Lending interest rate (%)	Portfolio Investment, net (BoP, current US$)	Net capital account (BoP, current US$)	Gross capital formation (current LCU)	GDP (constant LCU)	Exogenous spending	Compensation of employees (current LCU)	Inflation, consumer prices (annual %)	Capital-labor ratio	Previous Final consumption expenditure (constant LCU)]	Previous Portfolio Investment, net (BoP, current US$)	Previous Net capital account (BoP, current US$)	Previous Compensation of employees (current LCU)	Previous Gross capital formation (current LCU)
Fragile and conflict affected situations	0.52	0.52	0.66	0.70	0.72	0.53	0.57	0.55	0.59	0.81	0.40	0.72	0.79	0.59	0.39
France	0.00		0.36	0.86		0.43						0.33	1.00		
French Polynesia	0.50	0.27	0.90	0.97	0.75	0.68	0.38		0.61		0.53	0.92	0.93		0.75
Gabon		0.66		0.79	0.76	0.60	0.10		0.51		0.58		0.82		0.74
Gambia, The	0.64	0.81	0.86	0.24	0.71	0.51	0.42	0.50	0.37	0.22	0.31	0.83	0.29	0.75	0.65
Georgia	0.63	0.38	0.56	0.55	0.57	0.35	0.72	0.82	0.77	0.78	0.65	0.83	0.48	0.76	0.60
Germany		0.82	0.00	0.15	0.92	0.89	0.19	0.80	0.74	0.40	0.41	0.00	0.18	0.84	0.94
Ghana	0.80	0.59	0.88	0.43	0.65	0.70	0.71	0.68	0.49	0.66	0.34	0.88	0.49	0.52	0.64
Greece						0.50									
Greenland	0.50	0.26	0.95	0.56	0.55	0.35		0.83	0.35	0.47		1.00	0.49	0.72	0.52
Grenada	0.62														
Guam	0.35	0.76	0.60	0.38	0.75	0.76	0.75	0.62	0.59	0.46	0.51	0.55	0.44	0.70	0.65
Guatemala		0.46	0.60	0.77	0.83	0.46	0.09		1.00		0.59	0.73	0.71		0.82
Guinea		0.15	0.62	0.92	0.62	0.48			0.43		0.45	0.41	0.86		0.61
Guinea-Bissau		0.63	0.80	0.82	0.91	0.46			0.24			0.72	0.86		0.91
Guyana	0.62								0.70						
Haiti		0.50		0.53	0.98	0.78			0.63		0.46		0.43		0.97
Heavily indebted poor countries (HIPC)									0.50						
High income	0.54	0.62	0.97	0.90	0.89	0.67	0.75	0.69	0.36	0.59	0.58	0.97	0.82	0.52	0.78
High income: nonOECD	0.40	0.55	0.49	0.65	0.64	0.86	0.44	0.03	0.49	0.50	0.39	0.39	0.82	0.16	0.59
High income: OECD	0.62	0.61	0.48	0.61	0.40	0.36	0.56	0.62	0.34	0.97	0.72	0.51	0.80	0.69	0.47
Honduras	0.46	0.66	0.69	0.77	0.74	0.81	0.63	0.88	0.69	0.74	0.30	0.79	0.67	0.75	0.72
Hong Kong SAR, China	0.55	0.36	0.38	0.03	0.84	0.83	0.78	0.73	0.46	0.97	0.66	0.44	0.25	0.74	0.86
Hungary	0.85	0.51	0.92	0.58	0.89	0.63	0.67	0.88	0.45	0.66	0.58	0.86	0.51	0.85	0.90
Iceland	0.00	0.61			0.88	0.46	0.59	0.67	0.79	0.90	0.52			0.70	0.93
India		0.68	0.00	0.86	0.64	0.77			0.46			0.00	0.86		0.52
Indonesia	0.71	0.35	0.49	0.82	0.60	0.69	0.79	0.61	0.43	0.10	0.59	0.50	0.69	0.66	0.58
Iran						0.59									
Iraq	0.69	0.92	0.82	0.76	0.79	0.77		0.87	0.65	0.55	0.71	0.79	0.82	0.88	0.73
Ireland	0.50	0.48	0.64	0.72	0.69	0.73	0.69	0.64	0.59	0.93	0.51	0.74	0.69	0.64	0.61
Isle of Man	0.67	0.77	0.57	0.67	0.67	0.36	0.69	0.90	0.61	0.29	0.77	0.55	0.71	0.85	0.68
Israel	0.38	0.68	0.55	0.70	0.22	0.85	0.41		0.67		0.53	0.53	0.65		0.26
Italy	0.66	0.43	0.37	0.93	0.68	0.97	0.37	0.66	0.38	0.80	0.50	0.41	0.81	0.64	0.68
Jamaica	0.21		0.24	0.98	0.59	0.56	0.71	0.45	0.32	0.06	0.62	0.25	0.94	0.46	0.70
Japan		0.66	0.98	0.18	0.83	0.78	0.81	0.74	0.70	0.82	0.57	0.96	0.21	0.80	0.85
Jordan	0.00		0.28	0.45	0.56	0.55						0.26	0.29		0.55
Kazakhstan		0.77	0.71	0.94	0.33	0.41	0.27		0.40		0.31	0.40	0.85		0.36
Kenya		0.64	0.74	0.53	0.77	0.78	0.00	0.66	0.46	0.92	0.08	0.80	0.50	0.58	0.84
Kiribati	0.83								0.68						
Korea, Rep.	0.38	0.58	0.34	0.86	0.90	0.32	0.52	0.76	0.29	0.99	0.22	0.33	0.83	0.67	0.91
Kosovo	0.00	0.11	0.94		0.92	0.71	0.24	0.07	0.47	0.54	0.69	1.00		0.24	0.71
Kuwait	0.55	0.90	0.42	0.59	0.36	0.37	0.74	0.78	0.25	0.40	0.47	0.43	0.43	0.61	0.69
Kyrgyz Republic		0.55	0.81	0.79	0.85	0.55	0.24	0.40	0.00	0.72	0.73	0.72	0.78	0.42	0.89
Lao PDR		0.38	0.44	0.80	0.83	0.65	0.72	0.46	0.57	0.54	0.85	0.45	0.90	0.40	0.82

Country or Group	Wage and salaried workers, total (% of total employed)	Lending interest rate (%)	Portfolio Investment, net (BoP, current US$)	Net capital account (BoP, current US$)	Gross capital formation (current LCU)	GDP (constant LCU)	Exogenous spending	Compensation of employees (current LCU)	Inflation, consumer prices (annual %)	Capital-labor ratio	Previous Final consumption expenditure (constant LCU)]	Previous Portfolio Investment, net (BoP, current US$)	Previous Net capital account (BoP, current US$)	Previous Compensation of employees (current LCU)	Previous Gross capital formation (current LCU)
Latin America & Caribbean (all income levels)		0.55	0.00	0.65	0.53	0.35	0.66		0.24		0.47	0.00	0.70		0.50
Latin America & Caribbean (developing only)		0.78	0.90		0.86	0.69			0.55			0.96			0.89
Latvia						0.55									
Least developed countries: UN classification	0.54	0.87	0.55	0.57	0.48	0.54	0.71	0.83	0.48	0.65	0.68	0.65	0.58	0.68	0.39
Lebanon	0.57	0.05	0.39	0.49	0.79	0.73	0.94	0.85	0.64	0.48	0.58	0.53	0.46	0.85	0.65
Lesotho	0.42	0.45	0.85	0.85	0.75	0.75	0.92	0.63	0.32	0.66	0.59	0.86	0.75	0.66	0.74
Liberia	0.70	0.98	0.73	0.59	0.73	0.45	0.52	0.47	0.20	0.30	0.43	0.58	0.61	0.32	0.74
Libya	0.04	0.74	0.50	0.97	0.86	0.65	0.39	0.82	0.47	0.72	0.66	0.71	0.95	0.81	0.87
Liechtenstein		0.55	0.95	0.90	0.81	0.75	0.11		0.71		0.51	1.00	0.95		0.78
Lithuania	0.78	0.55	0.82	0.62	0.78	0.66	0.89	0.87	0.43	0.17	0.89	0.86	0.60	0.80	0.74
Low & middle income	0.06								0.69						
Low income	0.07	0.29	0.42	0.19	0.81	0.26	0.69	0.79	0.63	0.18	0.38	0.68	0.51	0.81	0.75
Lower middle income									0.39						
Luxembourg	0.00	0.16	0.88	0.94	0.89	0.66	0.59	0.56	0.65	0.88	0.60	0.85	0.92	0.65	0.70
Macao SAR, China	0.82	0.71	0.87	0.48	0.64	0.66	0.50	0.60	0.71	0.26	0.49	0.87	0.49	0.57	0.59
Macedonia, FYR			0.06	0.55		0.59						0.11	0.56		
Madagascar									0.45						
Malawi		0.26	0.14	0.80	0.91	0.66	0.63		0.20		0.43	0.05	0.50		0.90
Malaysia	0.97	0.66	0.64		0.65	0.44	0.61	0.80	0.39	0.38	0.44	0.85		0.81	0.62
Maldives	0.69	0.94	0.77		0.79	0.37	0.76	0.78	0.39	0.54	0.56	0.73		0.83	0.74
Mali		0.75	0.91	0.68		0.79						0.83	0.42		
Malta	0.38	0.50	0.57	0.86	0.61	0.45	0.77	0.53	0.54	0.09	0.05	0.69	0.89	0.56	0.63
Marshall Islands						0.44									
Mauritania	0.47	0.82	0.86	0.11	0.90	0.82	0.73	0.90	0.36	0.71	0.54	0.80	0.05	0.86	0.88
Mauritius	0.45	0.65	0.37	0.57	0.29	0.82	0.46		0.25		0.56	0.35	0.32		0.17
Mexico	0.45	0.52	0.64		0.75	0.68	0.72	0.85	0.63	0.63	0.46	0.53		0.86	0.74
Micronesia, Fed. Sts.		0.20	0.65	0.88	0.91	0.59	0.88	0.00	0.51	0.00	0.74	0.94	0.95	0.00	0.93
Middle East & North Africa (all income levels)		0.35		0.83	0.87	0.92	0.48		0.60		0.62		0.64		0.89
Middle East & North Africa (developing only)	0.17								0.53						
Middle income	0.50	0.35	0.55	0.70	0.85	0.71	0.85	0.70	0.39	0.69	0.68	0.59	0.68	0.65	0.86
Moldova		0.48		0.57	0.93	0.73	0.76	0.45	0.49	0.64	0.58		0.13	0.00	0.88
Monaco	0.80	0.41	0.83	0.91	0.85	0.65	0.78	0.79	0.56	0.56	0.55	0.87	0.97	0.75	0.72
Mongolia	0.74								0.49						
Montenegro			0.55	0.16	0.92	0.28						0.32	0.38		0.79
Morocco	0.58	0.29	0.64	0.80	0.72	0.68	0.33	0.87	0.71	0.42	0.60	0.49	0.74	0.79	0.55
Mozambique	0.11								0.53						
Myanmar	0.18								0.61						
Namibia	0.42	0.98	0.64	0.81	0.84	0.45	0.54	0.67	0.49	0.51	0.44	0.58	0.84	0.69	0.80
Nepal		0.13	0.02	0.49	0.90	0.84	0.00		0.63		0.22	0.07	0.11		0.90

Country or Group	Wage and salaried workers, total (% of total employed)	Lending interest rate (%)	Portfolio Investment, net (BoP, current US$)	Net capital account (BoP, current US$)	Gross capital formation (current LCU)	GDP (constant LCU)	Exogenous spending	Compensation of employees (current LCU)	Inflation, consumer prices (annual %)	Capital-labor ratio	Previous Final consumption expenditure (constant LCU)]	Previous Portfolio Investment, net (BoP, current US$)	Previous Net capital account (BoP, current US$)	Previous Compensation of employees (current LCU)	Previous Gross capital formation (current LCU)
Netherlands		0.42	0.86	0.94	0.87	0.83	0.19	0.66	0.55	0.75	0.76	0.88	0.72	0.67	0.86
New Caledonia															
New Zealand															
Nicaragua	0.82	0.40	0.89	0.96	0.66	0.87	0.45	0.91	0.68	0.59	0.43	0.85	0.93	0.98	0.63
Niger	0.95	0.58	0.71	0.81	0.78	0.59	0.49	0.68	0.62	0.60	0.62	0.72	0.87	0.66	0.78
Nigeria									0.22						
North America	0.54		0.91	0.94	0.84	0.84	0.61	0.78	0.44	0.62	0.49	0.81	0.97	0.64	0.86
North Korea						0.56									
Northern Mariana Islands	0.52	0.34	0.76	0.77	0.80	0.79	0.35	0.65	0.69	0.40	0.71	0.87	0.70	0.63	0.83
Norway		0.58	0.77	0.49	0.85	0.59	0.55	0.46	0.44	0.46	0.58	0.70	0.43	0.56	0.70
Not classified	0.58	0.55	0.87	0.92	0.67	0.80	0.29	0.39	0.59	0.58	0.78	0.86	0.93	0.26	0.72
OECD members									0.44						
Oman	0.64	0.96	0.61	0.73	0.91	0.65	0.55	0.70	0.46	0.89	0.62	0.59	0.67	0.65	0.93
Other small states	0.68	0.42	0.55	0.48	0.79	0.83	0.35	0.62	0.49	0.74	0.55	0.67	0.39	0.71	0.78
Pacific island small states	0.66	0.97	0.61	0.67	0.33	0.45	0.31	0.97	0.53	0.52	0.14	0.78	0.60	0.86	0.23
Pakistan	0.59	0.43	0.76	0.54	0.71	0.79	0.61	0.80	0.58	0.38	0.27	0.80	0.48	0.77	0.45
Palau	0.62				0.57	0.75	0.75				0.21				0.57
Panama	0.70	0.70	0.55	0.68	0.70	0.72		0.38	0.81	0.02	0.45	0.68	0.41	0.02	0.62
Papua New Guinea		0.13	0.70	0.66	0.87	0.30	0.62	0.55	0.17	0.87	0.76	0.61	0.80	0.61	0.87
Paraguay	0.65	0.69	0.77	0.84	0.67	0.76	0.47	0.84	0.38	0.67	0.50	0.81	0.83	0.77	0.65
Peru	0.55	0.77	0.46	0.69	0.82	0.19	0.78	0.60	0.67	0.60	0.76	0.73	0.70	0.67	0.84
Philippines	0.50	0.42	0.66	0.98	0.97	0.75	0.84	0.52	0.30	0.00	0.71	0.73	0.94	0.64	0.95
Poland	0.69								0.49						
Portugal	0.81	0.10	0.84	0.48		0.72		0.00	0.49			0.86	0.56	0.00	
Puerto Rico	0.77	0.60				0.68		0.95	0.67					0.94	
Qatar		0.59	0.93	0.99		0.20		0.69	0.64			0.78	0.97	0.67	
Romania	0.00		0.35	0.67	0.87	0.54			0.34		0.05	0.38	0.69		0.87
Russian Federation	0.07	0.18	0.83	0.91	0.76	0.63	0.70	0.90	0.48	0.86	0.63	0.44	0.50	0.88	0.74
Rwanda	0.37	0.83	0.23	0.83	0.56	0.43		0.38	0.30	0.23	0.46	0.07	0.91	0.29	0.41
Samoa									0.57						
San Marino		0.29	0.91	0.90	0.80	0.49	0.66	0.65	0.55	0.09	0.48	0.96	1.00	0.65	0.81
Sao Tome and Principe		0.75	0.76	0.87	0.98	0.78	0.84		0.53		0.68	0.56	0.86		0.95
Saudi Arabia	0.34	0.82	0.71	0.27	0.95	0.63	0.63	0.79	0.37	0.96	0.50	0.75	0.14	0.70	0.94
Senegal															
Serbia	0.60	0.33	0.51	0.66	0.50	0.48	0.70	0.81	0.57	0.63	0.40	0.64	0.57	0.83	0.25
Seychelles	0.52	0.91	0.39	0.55	0.72	0.54	0.80	0.88	0.71	0.40	0.50	0.34	0.58	0.41	0.72
Sierra Leone		0.58	0.94	0.73	0.91	0.29			0.55			0.91	0.74		0.88
Singapore					0.90	0.55	0.21				0.54				0.93
Sint Maarten (Dutch part)	0.92	0.43	0.71	0.79	0.80	0.80	0.60	0.87	0.55	0.29	0.44	0.80	0.94	0.85	0.81
Slovak Republic	0.67	0.60	0.73	0.93	0.87	0.50	0.67	0.72	0.58	0.76	0.57	0.82	0.97	0.70	0.65
Slovenia					0.32	0.10	0.40		0.07		0.02				0.32
Small states									0.44						
Solomon Islands									0.52						
Somalia	0.85	0.50	0.82	0.58	0.61	0.86	0.68	0.65	0.53	0.44	0.32	0.85	0.62	0.62	0.55

Country or Group	Wage and salaried workers, total (% of total employed)	Lending interest rate (%)	Portfolio Investment, net (BoP, current US$)	Net capital account (BoP, current US$)	Gross capital formation (current LCU)	GDP (constant LCU)	Exogenous spending	Compensation of employees (current LCU)	Inflation, consumer prices (annual %)	Capital-labor ratio	Previous Final consumption expenditure (constant LCU)]	Previous Portfolio Investment, net (BoP, current US$)	Previous Net capital account (BoP, current US$)	Previous Compensation of employees (current LCU)	Previous Gross capital formation (current LCU)
South Africa	0.47	0.22	0.76	0.39	0.86	0.72		0.60	0.55	0.31		0.67	0.43	0.59	0.85
South Asia	0.42								0.59						
South Sudan									0.41						
Spain		0.22	0.52	0.86	0.68	0.49		0.21	0.72	0.41		0.16	0.81	0.28	0.42
Sri Lanka									0.49						
St. Kitts and Nevis	0.47	0.37	0.82	0.64	0.69	0.56		0.38	0.28	0.31		0.68	0.73	0.31	0.56
St. Lucia	0.45								0.48						
St. Martin (French part)									0.30						
St. Vincent and the Grenadines															
Sub-Saharan Africa (all income levels)	0.75	0.25	0.72	0.71	0.48	0.66			0.66			0.72	0.81		0.48
Sub-Saharan Africa (developing only)			0.33	0.56	0.80	0.80	0.92		0.54		0.58	0.37	0.80		0.82
Sudan	0.96	0.51	0.58	0.56	0.88	0.78		0.49	0.50	0.10		0.68	0.71	0.49	0.92
Suriname		0.47	0.67	0.67	0.79	0.72	0.64		0.49		0.67	0.68	0.73		0.79
Swaziland	0.63	0.56	0.80	0.39	0.72	0.91	0.89	0.77	0.53	0.54	0.60	0.74	0.37	0.86	0.48
Sweden	0.45	0.60	0.53	0.75	0.62	0.50	0.82	0.60	0.58	0.85	0.48	0.84	0.66	0.45	0.54
Switzerland	0.40	0.80	0.11	0.75	0.89	0.55	0.46	0.56	0.62	0.77	0.40	0.06	0.90	0.58	0.92
Syrian Arab Republic		0.94	0.77	0.64	0.74	0.29	0.30	0.00	0.48	0.56	0.27	0.82	0.46	0.00	0.85
Tajikistan	0.21	0.61	0.55	0.97	0.65	0.85	0.55	0.05	0.64	0.55	0.54	0.54	1.00	0.00	0.57
Tanzania	0.79	0.49	0.58	0.05	0.93	0.69	0.68	0.45	0.56	0.94	0.59	0.54	0.05	0.65	0.70
Thailand		0.41	0.10	0.30	0.53	0.53	0.45		0.70		0.31	0.08	0.15		0.68
Timor-Leste		0.10	0.32	0.85	0.78	0.50	0.98	0.86	0.47	0.83	0.54	0.33	0.76	0.84	0.77
Togo		0.29	0.61	0.61	0.65	0.45			0.55			0.76	0.70		0.67
Tonga	0.19	0.50	0.86		0.76	0.71	0.74	0.54	0.52	0.11	0.48	0.85		0.71	0.74
Trinidad and Tobago	0.65	0.55	0.69	0.77	0.67	0.73	0.50	0.61	0.40	0.43	0.73	0.63	0.70	0.59	0.67
Tunisia	0.58		0.86	0.50	0.85	0.81	0.75	0.89	0.47	0.78	0.51	0.84	0.19	0.91	0.86
Turkey					0.88	0.79									0.87
Turkmenistan	0.05														
Turks and Caicos Islands			0.75	0.06		0.30						0.43	0.11		
Tuvalu	0.04	0.25	0.13	0.95	0.88	0.63	0.54	0.80	0.58	0.88	0.60	0.26	0.97	0.63	0.87
Uganda	0.54	0.77	0.55	0.83	0.71	0.26	0.68	0.77	0.51	0.56	0.59	0.59	0.86	0.74	0.73
Ukraine	0.00								0.50						
United Arab Emirates															
United Kingdom	0.67				0.12	0.77		0.03	0.44	0.00	0.07			0.06	0.14
United States	0.62	0.54	0.76	0.64	0.66	0.85	0.36	0.84	0.56	0.65	0.51	0.89	0.35	0.80	0.57
Upper middle income	0.70	0.73	0.77	0.60	0.71	0.83	0.70	0.65	0.51	0.82	0.69	0.80	0.37	0.53	0.54
Uruguay	0.38	0.45	0.79	0.89	0.82	0.61	0.70	0.58	0.28	0.66	0.59	0.70	0.88	0.55	0.83
Uzbekistan					0.75	0.66									0.78
Vanuatu		0.29	0.54	0.54	0.72	0.50	0.65	0.74	0.65	0.58	0.46	0.70	0.41	0.71	0.73
Venezuela, RB	0.49	0.65	0.66		0.92	0.74	0.83	0.74	0.48	0.17	0.60	0.83		0.82	0.94
Vietnam	0.50	0.92	0.63		0.76	0.85	0.40		0.12		0.72	0.93			0.72
Virgin Islands (U.S.)						0.42									
West Bank and Gaza									0.38						
World	0.47	0.35	0.65	0.72	0.60	0.58	0.34	0.29	0.54	0.19	0.24	0.85	0.74	0.11	0.58
Yemen, Rep.	0.04	0.71	0.54	0.18	0.35	0.33			0.78			0.55	0.14		0.37

Country or Group	Wage and salaried workers, total (% of total employed)	Lending interest rate (%)	Portfolio Investment, net (BoP, current US$)	Net capital account (BoP, current US$)	Gross capital formation (current LCU)	GDP (constant LCU)	Exogenous spending	Compensation of employees (current LCU)	Inflation, consumer prices (annual %)	Capital-labor ratio	Previous Final consumption expenditure (constant LCU)]	Previous Portfolio Investment, net (BoP, current US$)	Previous Net capital account (BoP, current US$)	Previous Compensation of employees (current LCU)	Previous Gross capital formation (current LCU)
Zambia	0.90	0.79	0.85	0.84		0.78	0.67	0.62	0.63		0.97	0.86	0.92	0.67	
Zimbabwe	0.57	0.94	0.53		0.70	0.37	0.65	0.00	0.69	0.59	0.31	0.72		0.05	0.46

6.2. Accuracy results for the models generated with the UNESCO Domain Knowledge model

Country name	Labor force with primary education (% of total)	Labor force with secondary education (% of total)	Labor force with tertiary education (% of total)	Scientific and technical journal articles	Trademark applications, total	General government final consumption expenditure (% of GDP)	Net official development assistance and official aid received	Agriculture, value added (% of GDP)	Industry, value added (% of GDP)	Manufacturing, value added (% of GDP)	Services, etc., value added (% of GDP)	Unemployment, total (% of total labor force)	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG]	GDP per capita, PPP (constant 2011 international $)	Previous GDP growth (annual %)	Previous GDP per capita, PPP (constant 2011 international $)
Afghanistan				0.76		0.25	0.79	0.26	0.47	0.34	0.48	0.40	0.39	0.55	0.54	0.75
Albania	0.33	0.07	0.10	0.59	0.42	0.84	0.41	0.59	0.24	0.60	0.61	0.64	0.85	0.38	0.79	0.36
Albania				0.60	0.70	0.50	0.42	0.57	0.44	0.11	0.08	0.78	0.46	0.54	0.68	0.52
Algeria				0.65	0.88	0.64	0.90	0.48	0.55	0.32	0.41	0.42	0.37	0.49	0.91	0.39
Algeria				0.69	0.82	0.54	0.61	0.54	0.75	0.66	0.69	0.32	0.83	0.51	0.86	0.65
Algeria				0.66	0.75	0.51	0.69	0.48	0.54	0.29	0.46	0.27	0.66	0.58	0.67	0.56
Algeria				0.13	0.61	0.70	0.58	0.34	0.39	0.40	0.45	0.31	0.42	0.55	0.24	0.52
Algeria	0.81	0.77	0.78	0.58	0.78	0.56	0.69	0.35	0.32	0.27	0.53	0.73	0.63	0.44	0.61	0.65
American Samoa																
Andorra				0.80	0.93			0.51	0.61	0.05	0.61		0.63		0.65	
Angola				0.31		0.54	0.89	0.78	0.39	0.77	0.38	0.57	0.82		0.76	
Antigua and Barbuda				0.83	0.41	0.47	0.54	0.83	0.32	0.70	0.38		0.43	0.63	0.31	0.58

Country name	Labor force with primary education (% of total)	Labor force with secondary education (% of total)	Labor force with tertiary education (% of total)	Scientific and technical journal articles	Trademark applications, total	General government final consumption expenditure (% of GDP)	Net official development assistance and official aid received	Agriculture, value added (% of GDP)	Industry, value added (% of GDP)	Manufacturing, value added (% of GDP)	Services, etc., value added (% of GDP)	Unemployment, total (% of total labor force)	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG]	GDP per capita, PPP (constant 2011 international $)	Previous GDP growth (annual %)	Previous GDP per capita, PPP (constant 2011 international $)
Argentina	0.75	0.51	0.54	0.37	0.57	0.38	0.87	0.32	0.48	0.44	0.41	0.43	0.27		0.28	
Armenia	0.04	0.82	0.82	0.92	0.39	0.71	0.64	0.71	0.21	0.60	0.66	0.73	0.82	0.55	0.85	0.60
Aruba					0.17	0.15	0.57	0.49	0.62	0.46	0.62		0.40		0.50	
Australia	0.69	0.85	0.61	0.58	0.68	0.76		0.68	0.56	0.72	0.59	0.53	0.49	0.48	0.69	0.37
Austria	0.58	0.83	0.67	0.61	0.70	0.49		0.59	0.45	0.55	0.70	0.24	0.50	0.51	0.46	0.55
Azerbaijan	0.06	0.00	0.06	0.63	0.59	0.82	0.62	0.44	0.60	0.66	0.55	0.58	0.44	0.61	0.66	0.57
Bahamas	0.10	0.57	0.64	0.28	0.58	0.40	0.90	0.66	0.56	0.47	0.23	0.32	0.46	0.35	0.78	0.19
Bahrain	0.59	0.05	0.73	0.53	0.80	0.39	0.80	0.53	0.72	0.58	0.72	0.86	0.69	0.79	0.63	0.75
Bangladesh				0.59	0.64	0.75	0.56	0.50	0.73	0.76	0.62	0.41	0.81	0.65	0.73	0.72
Barbados				0.48	0.56	0.47	0.52					0.81	0.22	0.45	0.22	0.27
Belarus				0.71	0.16	0.33	0.77	0.53	0.13	0.53	0.76	0.66	0.55	0.58	0.46	0.60
Belgium	0.52	0.14	0.19	0.66	0.75	0.54		0.34	0.48	0.52	0.36	0.47	0.50	0.53	0.49	0.41
Belize	0.60	0.85	0.64	0.27	0.64	0.79	0.26	0.66	0.61	0.57	0.76	0.63	0.52	0.69	0.48	0.63
Benin				0.28		0.44	0.40	0.58	0.70	0.68	0.56	0.24	0.39	0.25	0.58	0.24
Bermuda	0.65	0.42	0.58			0.14	0.98	0.24	0.28	0.21	0.19		0.55	0.44	0.67	0.18
Bhutan				0.63	0.18	0.49	0.37	0.25	0.46	0.35	0.24	0.50	0.68	0.45	0.76	0.52
Bolivia	0.17	0.56	0.85	0.34	0.68	0.65	0.69	0.36	0.50	0.82	0.71	0.40	0.74	0.58	0.77	0.54
Bosnia and Herzegovina	0.21	0.32	0.36	0.86	0.44	0.08	0.51	0.70	0.47	0.40	0.70	0.48	0.90	0.58	0.82	0.54
Botswana	0.00	0.25		0.24	0.49	0.51	0.83	0.74	0.51	0.77	0.40	0.18	0.75	0.64	0.73	0.66
Brazil	0.79	0.50	0.66	0.35	0.60	0.71	0.52	0.52	0.36	0.84	0.75	0.58	0.53	0.48	0.55	0.45
Brunei Darussalam				0.60	0.79	0.51	0.72	0.71	0.43	0.43	0.42	0.23	0.32	0.50	0.83	0.58
Bulgaria	0.52	0.54	0.37	0.41	0.85	0.68	0.14	0.46	0.58		0.75	0.29	0.51	0.49	0.45	0.73
Burkina Faso				0.32	0.16	0.50	0.40	0.42	0.58	0.52	0.40	0.23	0.35	0.42	0.30	0.45
Burundi				0.42	0.67	0.64	0.55	0.40	0.57	0.41	0.51	0.56	0.37	0.65	0.78	0.68
Cabo Verde				0.70		0.41	0.66	0.35	0.79	0.36	0.60	0.76	0.32	0.48	0.38	0.65
Cambodia	0.59	0.09	0.59	0.55	0.61	0.70	0.51	0.64	0.58	0.37	0.56	0.57	0.53	0.45	0.62	0.43
Cameroon				0.64		0.44	0.67	0.55	0.55	0.71	0.60	0.54	0.77	0.66	0.77	0.62
Canada	0.46	0.68	0.62	0.83	0.70	0.54		0.65	0.05	0.80	0.05	0.53	0.54	0.71	0.48	0.55
Canada	0.61	0.57	0.71	0.64	0.95	0.68	0.76	0.54	0.81	0.42	0.75	0.59	0.72	0.37	0.60	0.59
Cayman Islands	0.00	0.00	0.70				0.61	0.06	0.06	0.00						
Central African Republic				0.42		0.39	0.34	0.43	0.56	0.41	0.64	0.84	0.88	0.86	0.95	0.83
Chad				0.54		0.53	0.54	0.87	0.96	0.80	0.86	0.69	0.60	0.57	0.78	0.91
Channel Islands													0.13		0.11	
Chile	0.61	0.38	0.38	0.71	0.68	0.40	0.77	0.68	0.45	0.69	0.73	0.16	0.60	0.62	0.57	0.33
China				0.79	0.73	0.56	0.61	0.56	0.75	0.57	0.67	0.67	0.82	0.61	0.90	0.59
Colombia	0.43	0.73	0.05	0.50	0.79	0.87	0.52	0.55	0.34	0.69	0.80	0.20	0.51	0.57	0.37	0.59
Comoros				0.73		0.45	0.56	0.60	0.61	0.65	0.36	0.67	0.63	0.48	0.80	0.45
Costa Rica	0.13	0.67	0.71	0.46	0.76	0.38	0.80	0.79	0.63	0.76	0.85	0.41	0.55	0.53	0.63	0.75
Costa Rica				0.63	0.62	0.44	0.52	0.41	0.61	0.50	0.55	0.52	0.50	0.72	0.59	0.34
Côte d'Ivoire				0.42		0.43	0.78	0.65	0.48	0.56	0.40	0.80	0.59	0.38	0.56	0.36
Croatia	0.44	0.93	0.56	0.35	0.88	0.75	0.42	0.63	0.23	0.47	0.64	0.47	0.53	0.67	0.51	0.55
Cuba	0.39	0.07	0.00	0.51	0.68	0.73	0.46	0.62	0.74	0.72	0.65	0.62	0.51	0.56	0.74	0.55
Curaçao					0.93											
Cyprus	0.51	0.42	0.47	0.56	0.78	0.44	0.84	0.71	0.53	0.74	0.87	0.73	0.59	0.49	0.61	0.30
Czech Republic	0.87	0.76	0.58	0.79	0.66	0.56	0.53	0.52	0.37	0.16	0.36	0.56	0.64	0.29	0.78	0.40

Country name	Labor force with primary education (% of total)	Labor force with secondary education (% of total)	Labor force with tertiary education (% of total)	Scientific and technical journal articles	Trademark applications, total	General government final consumption expenditure (% of GDP)	Net official development assistance and official aid received	Agriculture, value added (% of GDP)	Industry, value added (% of GDP)	Manufacturing, value added (% of GDP)	Services, etc., value added (% of GDP)	Unemployment, total (% of total labor force)	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG]	GDP per capita, PPP (constant 2011 international $)	Previous GDP growth (annual %)	Previous GDP per capita, PPP (constant 2011 international $)
Democratic Republic of Congo				0.47	0.55	0.61	0.92	0.58	0.64	0.31	0.34		0.78	0.89	0.69	0.81
Denmark	0.32	0.75	0.68	0.60	0.61	0.65		0.48	0.59	0.83	0.59	0.50	0.55	0.70	0.66	0.64
Djibouti				0.71		0.65	0.60	0.60	0.73	0.88	0.71		0.61	0.55	0.56	0.52
Dominica				0.46	0.29	0.63	0.62	0.50	0.51	0.21	0.48		0.50	0.20	0.61	0.19
Dominican Republic	0.93	0.54	0.43	0.74	0.45	0.68	0.42	0.50	0.65	0.57	0.65	0.31	0.44	0.19	0.73	0.33
Ecuador	0.50	0.54	0.42	0.48	0.69	0.55	0.31	0.38	0.65	0.37	0.46	0.47	0.67	0.63	0.71	0.73
Egypt	0.53	0.60	0.20	0.83	0.66	0.55	0.72	0.54	0.70	0.33	0.44	0.54	0.66	0.66	0.47	0.46
Egypt	0.84	0.82	0.82	0.68	0.81	0.83	0.72	0.59	0.57	0.75	0.91	0.17	0.52	0.78	0.55	0.78
El Salvador	0.76	0.17	0.10	0.18	0.57	0.61	0.56	0.28	0.19	0.22	0.66	0.75	0.86	0.68	0.79	0.30
Equatorial Guinea				0.75		0.90	0.45					0.37	0.91	0.27	0.94	0.28
Eritrea				0.54		0.45	0.68	0.44	0.56	0.63	0.18	0.40	0.57	0.53	0.69	0.50
Estonia	0.36	0.83	0.54	0.35	0.66	0.57	0.52	0.66	0.51	0.41	0.65	0.58	0.55	0.86	0.73	0.61
Ethiopia	0.83	0.38	0.83	0.46	0.54	0.38	0.60	0.44	0.55	0.36	0.41	0.54	0.45	0.74	0.55	0.72
Ethiopia	0.89	0.89	0.80	0.73	0.88	0.69	0.71	0.73	0.53	0.78	0.79	0.48	0.66	0.79	0.58	0.71
Ethiopia				0.28		0.40	0.68	0.49	0.18	0.81	0.40	0.22	0.63	0.54	0.48	0.48
Faroe Islands						0.09				0.38	0.23					
Fiji				0.51	0.57	0.40	0.42	0.59	0.50	0.55	0.51	0.40	0.51	0.13	0.51	0.33
Finland	0.53	0.67	0.63	0.59	0.74	0.54		0.62	0.38	0.74	0.81	0.58	0.52	0.55	0.52	0.38
France	0.75	0.66	0.78	0.63	0.82	0.80		0.59	0.58	0.69	0.87	0.15	0.70	0.61	0.48	0.50
French Polynesia	0.50	0.06	0.00			0.44	0.37	0.48					0.27		0.35	
Gabon				0.16		0.64	0.39	0.37	0.05	0.15	0.37	0.70	0.53	0.24	0.82	0.31
Gambia				0.30	0.15	0.71	0.55	0.31	0.43	0.50	0.13	0.68	0.56	0.51	0.53	0.54
Georgia	0.87	0.80	0.98	0.46	0.52	0.53	0.30	0.74	0.49	0.73	0.47	0.29	0.86	0.48	0.90	0.46
Germany	0.48	0.39	0.43	0.78	0.89	0.52		0.42	0.72	0.55	0.66	0.50	0.44	0.13	0.52	0.31
Ghana	0.80	0.00	0.55	0.68	0.52	0.59	0.70	0.61	0.56	0.57	0.54	0.29	0.39	0.68	0.66	0.77
Greece	0.66	0.60	0.78	0.78	0.39	0.81		0.52	0.59	0.55	0.54	0.85	0.78	0.60	0.54	0.58
Greenland													0.27		0.24	
Grenada				0.75		0.48	0.69	0.61	0.49	0.57	0.53		0.54	0.64	0.56	0.67
Guam																
Guatemala	0.83	0.40	0.15	0.34	0.31	0.57	0.45	0.65	0.39	0.78	0.23	0.34	0.59	0.27	0.74	0.20
Guinea				0.78		0.59	0.53	0.25	0.39	0.73	0.53	0.79	0.49	0.78	0.56	0.58
Guinea-Bissau				0.43	0.67	0.37	0.58	0.42	0.54	0.68	0.38	0.40	0.55	0.65	0.63	0.68
Guyana	0.05	0.09	0.00	0.53	0.02	0.45	0.68	0.56	0.55	0.63	0.55	0.59	0.59	0.56	0.44	0.59
Guyana	0.67	0.51	0.75	0.46	0.65	0.63	0.76	0.65	0.47	0.18	0.48	0.69	0.72	0.74	0.45	0.22
Haiti				0.81	0.43	0.37	0.92					0.60	0.48	0.14	0.28	0.12
Honduras				0.67	0.66	0.55	0.62	0.41	0.67	0.67	0.63	0.68	0.57	0.64	0.64	0.73
Honduras				0.73		0.42	0.29	0.77	0.54	0.58	0.39	0.74	0.62	0.53	0.42	0.57
Hong Kong	0.14	0.64	0.51	0.38	0.51	0.62	0.68	0.27	0.60	0.34	0.60	0.48	0.45	0.86	0.60	0.60
Hungary	0.68	0.59	0.53	0.44	0.63	0.70	0.66	0.54	0.36	0.48	0.45	0.33	0.48	0.53	0.60	0.38
Iceland	0.40	0.81	0.23	0.42	0.89	0.60		0.04	0.19	0.40	0.59	0.58	0.53	0.48	0.46	0.28
India	0.05	0.05	0.23	0.76	0.73	0.52	0.48	0.42	0.59	0.60	0.57	0.14	0.59	0.51	0.54	0.65
India																
Indonesia	0.92	0.73	0.71	0.64	0.81	0.58	0.60	0.65	0.73	0.83	0.78	0.53	0.77	0.41	0.80	0.65
Iran	0.75	0.75	0.00	0.66	0.84	0.60	0.89	0.68	0.55	0.45	0.45	0.46	0.43	0.56	0.56	0.41

Country name	Labor force with primary education (% of total)	Labor force with secondary education (% of total)	Labor force with tertiary education (% of total)	Scientific and technical journal articles	Trademark applications, total	General government final consumption expenditure (% of GDP)	Net official development assistance and official aid received	Agriculture, value added (% of GDP)	Industry, value added (% of GDP)	Manufacturing, value added (% of GDP)	Services, etc., value added (% of GDP)	Unemployment, total (% of total labor force)	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG]	GDP per capita, PPP (constant 2011 international $)	Previous GDP growth (annual %)	Previous GDP per capita, PPP (constant 2011 international $)
Iraq				0.80	0.70	0.62	0.92					0.96	0.71	0.46	0.78	0.55
Ireland	0.32	0.74	0.43	0.55	0.43	0.42		0.72	0.20	0.43	0.54	0.32	0.56	0.59	0.52	0.52
Isle of Man													0.66		0.68	
Israel	0.50	0.76	0.40	0.71	0.67	0.35	0.56					0.54	0.74	0.57	0.69	0.42
Italy	0.96	0.93	0.75	0.71	0.95	0.71		0.83	0.80	0.85	0.85	0.24	0.73	0.27	0.63	0.46
Jamaica				0.57	0.57	0.49	0.56	0.46	0.76	0.75	0.71	0.78	0.51	0.68	0.50	0.38
Japan	0.59	0.22	0.76	0.60	0.47	0.37		0.57	0.20	0.35	0.57	0.28	0.58	0.14	0.56	0.25
Jordan	0.23	0.30	0.22	0.81	0.77	0.57	0.78	0.62	0.78	0.71	0.54	0.59	0.61	0.54	0.85	0.59
Kazakhstan	0.65	0.17	0.75	0.71	0.68	0.79	0.46	0.82	0.65	0.60	0.90	0.50	0.51	0.62	0.46	0.81
Kenya				0.36	0.69	0.30	0.73	0.58	0.42	0.53	0.37	0.34	0.70	0.58	0.72	0.62
Kiribati				0.90		0.59	0.58	0.79	0.93	0.52	0.89		0.78	0.71	0.91	0.64
Kosovo						0.51	0.05	0.14	0.57	0.38	0.41		0.92	0.59	0.94	0.48
Kuwait	0.58	0.38	0.00	0.63	0.63	0.87	0.76	0.20	0.55	0.30	0.75	0.69	0.61	0.54	0.74	0.65
Kuwait	0.08	0.74	0.68	0.45	0.70	0.68	0.58	0.50	0.57	0.58	0.81	0.33	0.58	0.46	0.64	0.46
Kyrgyz Republic	0.50	0.00	0.00	0.57	0.39	0.63	0.82	0.55	0.45	0.52	0.24	0.83	0.76	0.34	0.76	0.41
Laos				0.66	0.98	0.60	0.38	0.66	0.56	0.59	0.59	0.57	0.61	0.57	0.88	0.55
Latvia	0.47	0.57	0.61	0.66	0.70	0.42	0.17	0.95	0.80	0.58	0.74	0.51	0.70	0.42	0.71	0.37
Lebanon				0.71	0.86	0.66	0.72	0.49	0.16	0.37	0.88	0.64	0.52	0.54	0.85	0.54
Lesotho				0.22	0.30	0.67	0.40	0.66	0.62	0.69	0.63	0.51	0.36	0.50	0.85	0.40
Liberia				0.76	0.13	0.63	0.87					0.90	0.64	0.49	0.65	0.63
Libya				0.35	0.76	0.18	0.90	0.68	0.82	0.77	0.91	0.76	0.42	0.63	0.70	0.10
Liechtenstein				0.49	0.65				0.54				0.50		0.51	
Lithuania	0.39	0.67	0.43	0.62	0.56	0.54	0.33	0.51	0.56	0.29	0.64	0.50	0.79	0.52	0.86	0.45
Luxembourg	0.55	0.51	0.78	0.78	0.06	0.56		0.52	0.66	0.57	0.61	0.53	0.58	0.61	0.62	0.51
Macao	0.57	0.43	0.52		0.69	0.37	0.87		0.45	0.69	0.44	0.53	0.24	0.61	0.36	0.72
Macedonia	0.45	0.39	0.10	0.23	0.70	0.60	0.39	0.65	0.54	0.54	0.61	0.49	0.60	0.60	0.52	0.58
Madagascar	0.00	0.61	0.61	0.30	0.71	0.50	0.67	0.31	0.68	0.56	0.33	0.27	0.54	0.55	0.38	0.36
Malawi				0.57	0.52	0.79	0.48	0.49	0.70	0.34	0.67	0.44	0.62	0.21	0.67	0.32
Malaysia	0.75	0.75	0.27	0.84	0.60	0.53	0.79	0.56	0.64	0.67	0.76	0.39	0.77	0.31	0.76	0.37
Malaysia	0.48	0.52	0.02	0.86	0.63	0.49	0.97	0.49	0.54	0.61	0.64	0.54	0.50	0.52	0.55	0.43
Maldives				0.38	0.55	0.73	0.63	0.60	0.47	0.32	0.30	0.31	0.79	0.49	0.64	0.54
Maldives				0.63	0.66	0.53	0.36	0.50	0.28	0.05	0.29	0.62	0.44	0.67	0.52	0.44
Mali				0.57	0.93	0.56	0.56	0.69	0.44	0.44	0.66	0.15	0.54	0.27	0.56	0.39
Malta	0.35	0.08	0.09	0.48	0.79	0.68	0.33	0.58	0.73	0.69	0.85	0.28	0.53	0.65	0.52	0.64
Marshall Islands				0.83			0.68	0.45	0.00	0.33	0.18		0.55	0.43	0.49	0.26
Marshall Islands				0.66	0.66	0.62	0.74	0.35	0.84	0.47	0.68	0.44	0.54	0.60	0.53	0.63
Mauritania				0.50		0.73	0.58	0.78	0.33	0.51	0.29	0.77	0.71	0.61	0.71	0.77
Mauritius	0.77	0.62	0.73	0.37	0.54	0.71	0.41	0.50	0.46	0.44	0.52	0.37	0.77	0.31	0.83	0.55
Mexico	0.49	0.65	0.62	0.61	0.69	0.43	0.50	0.44	0.51	0.51	0.39	0.62	0.63	0.13	0.62	0.15
Micronesia				0.76			0.68	0.88	0.70	0.50	0.26		0.52	0.41	0.43	0.38
Moldova	0.37	0.40	0.31	0.51	0.55	0.29	0.61	0.28	0.70	0.81	0.49	0.66	0.80	0.58	0.77	0.67
Monaco				0.33	0.64								0.75		0.74	
Mongolia	0.14	0.58	0.32	0.71	0.35	0.61	0.47	0.63	0.43	0.74	0.24	0.56	0.62	0.66	0.60	0.63
Montenegro	0.32	0.80	0.13	0.00	0.93	0.48	0.65	0.40	0.31	0.40	0.32	0.55	0.41	0.51	0.33	0.48
Morocco	0.16	0.81	0.06	0.63	0.58	0.46	0.64	0.27	0.60	0.18	0.63	0.45	0.57	0.47	0.60	0.58

Country name	Labor force with primary education (% of total)	Labor force with secondary education (% of total)	Labor force with tertiary education (% of total)	Scientific and technical journal articles	Trademark applications, total	General government final consumption expenditure (% of GDP)	Net official development assistance and official aid received	Agriculture, value added (% of GDP)	Industry, value added (% of GDP)	Manufacturing, value added (% of GDP)	Services, etc., value added (% of GDP)	Unemployment, total (% of total labor force)	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG]	GDP per capita, PPP (constant 2011 international $)	Previous GDP growth (annual %)	Previous GDP per capita, PPP (constant 2011 international $)
Mozambique				0.76	0.71	0.30	0.48	0.36	0.43	0.26	0.65	0.47	0.35	0.80	0.73	0.56
Myanmar				0.37	0.75		0.99	0.35	0.32	0.13	0.40	0.72	0.56		0.54	
Myanmar				0.69	0.59	0.62	0.61	0.72	0.49	0.56	0.71	0.68	0.66	0.48	0.66	0.28
Namibia	0.00	0.00	0.00	0.35	0.66	0.57	0.45	0.25	0.76	0.47	0.81	0.92	0.48	0.50	0.49	0.53
Nepal				0.38	0.85	0.61	0.26	0.45	0.67	0.74	0.79	0.45	0.40	0.54	0.72	0.40
Netherlands	0.43	0.53	0.39	0.39	0.49	0.66		0.60	0.49	0.58	0.73	0.14	0.73	0.51	0.68	0.52
Netherlands	0.11	0.81	0.81	0.72	0.34	0.27	0.48	0.71	0.56	0.27	0.76	0.50	0.31	0.56	0.28	0.56
New Caledonia						0.93	0.60	0.60	0.78	0.51	0.78		0.44		0.52	
New Zealand	0.63	0.76	0.30	0.45	0.77	0.53		0.67	0.48	0.73	0.61	0.24	0.45	0.46	0.54	0.33
New Zealand	0.55	0.64	0.43	0.53	0.80	0.67	0.69	0.65	0.71	0.44	0.72	0.51	0.64	0.71	0.42	0.50
New Zealand	0.36	0.35	0.57	0.42	0.69	0.62	0.58	0.48	0.51	0.26	0.41	0.43	0.59	0.52	0.37	0.38
Nicaragua	0.79	0.64	0.64	0.49	0.56	0.76	0.64	0.58	0.52	0.62	0.20	0.58	0.71	0.45	0.72	0.23
Niger				0.25		0.31	0.51	0.51	0.50	0.36	0.74		0.41	0.67	0.45	0.61
Nigeria				0.65	0.76	0.67	0.94	0.55	0.25	0.28	0.72	0.35	0.68	0.40	0.63	0.42
North Korea				0.71	0.82		0.64					0.78				
Northern Mariana Islands							0.47									
Norway	0.62	0.25	0.55	0.59	0.88	0.60		0.51	0.48	0.46	0.52	0.54	0.64	0.83	0.46	0.59
Oman				0.65	0.68	0.57	0.81	0.90	0.63	0.89	0.49	0.51	0.93	0.17	0.93	0.45
Oman				0.64	0.70	0.49	0.27	0.70	0.50	0.57	0.48	0.45	0.49	0.34	0.58	0.77
Pakistan	0.04	0.08	0.00	0.78	0.79	0.53	0.47	0.56	0.58	0.60	0.57	0.59	0.55	0.55	0.58	0.49
Palau				0.45			0.90	0.90	0.69	0.85	0.63		0.15	0.52	0.15	0.39
Panama	0.83	0.75	0.66	0.60	0.54	0.45	0.76	0.70	0.19	0.58	0.59	0.70	0.67	0.63	0.73	0.76
Papua New Guinea				0.69	0.61	0.70	0.51	0.50	0.59	0.43	0.67	0.40	0.40	0.55	0.47	0.54
Paraguay	0.08	0.45	0.50	0.63	0.33	0.69	0.54	0.58	0.42	0.43	0.39	0.52	0.64	0.77	0.62	0.81
Paraguay				0.44		0.84	0.47	0.46	0.48	0.57	0.35		0.41	0.65	0.43	0.32
Peru	0.78	0.82	0.52	0.59	0.66	0.44	0.44	0.48	0.64	0.40	0.51	0.29	0.58	0.63	0.49	0.72
Philippines	0.50	0.23	0.14	0.69	0.86	0.62	0.54	0.72	0.56	0.56	0.73	0.63	0.75	0.74	0.72	0.73
Poland	0.56	0.38	0.67	0.57	0.88	0.82	0.75	0.74	0.36	0.56	0.72	0.39	0.89	0.37	0.82	0.35
Portugal	0.49	0.71	0.60	0.63	0.81	0.77		0.68	0.31	0.49	0.28	0.75	0.61	0.65	0.57	0.65
Puerto Rico						0.58		0.49	0.65	0.58	0.65	0.32	0.29	0.56	0.34	0.53
Qatar				0.79	0.77	0.63	0.37	0.80	0.31	0.17	0.34	0.80	0.40	0.35	0.32	0.08
Republic of Congo				0.38		0.43	0.94	0.38	0.51	0.45	0.49	0.43	0.70	0.37	0.72	0.42
Romania	0.75	0.75	0.70	0.47	0.96	0.51	0.85	0.59	0.63	0.65	0.66	0.44	0.65	0.55	0.57	0.71
Russia	0.89	0.67	0.71	0.60	0.48	0.53	0.62	0.83	0.67	0.43	0.67	0.46	0.47	0.42	0.49	0.36
Rwanda				0.33	0.82	0.50	0.55	0.69	0.38	0.51	0.53		0.65	0.39	0.36	0.26
Rwanda	0.00	0.00	0.00	0.71	0.72	0.35	0.57	0.36	0.54	0.44	0.63	0.18	0.56	0.55	0.51	0.59
Samoa				0.49	0.80		0.43						0.27	0.34	0.30	0.27
San Marino	0.46	0.20	0.28	0.73	0.25								0.46		0.47	
São Tomé and Principe				0.78	0.51		0.34	0.57	0.55	0.39	0.39		0.63	0.66	0.69	0.39
Saudi Arabia	0.53	0.47	0.53	0.91		0.54	0.57	0.42	0.38	0.54	0.60	0.22	0.63	0.70	0.58	0.74
Senegal				0.70		0.75	0.24	0.52	0.47	0.46	0.20	0.69	0.28	0.25	0.35	0.29
Serbia	0.86	0.86	0.06	0.91	0.48	0.45	0.58	0.70	0.66	0.58	0.78	0.42	0.74	0.62	0.77	0.32
Serbia				0.69		0.32	0.41	0.81	0.63	0.76	0.57	0.42	0.45	0.58	0.40	0.71
Seychelles				0.45	0.61	0.58	0.52	0.65	0.67	0.61	0.52		0.49	0.28	0.51	0.28
Sierra Leone				0.54	0.60	0.49	0.60	0.44	0.55	0.69	0.61		0.55	0.68	0.80	0.67

Country name	Labor force with primary education (% of total)	Labor force with secondary education (% of total)	Labor force with tertiary education (% of total)	Scientific and technical journal articles	Trademark applications, total	General government final consumption expenditure (% of GDP)	Net official development assistance and official aid received	Agriculture, value added (% of GDP)	Industry, value added (% of GDP)	Manufacturing, value added (% of GDP)	Services, etc., value added (% of GDP)	Unemployment, total (% of total labor force)	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG]	GDP per capita, PPP (constant 2011 international $)	Previous GDP growth (annual %)	Previous GDP per capita, PPP (constant 2011 international $)
Singapore	0.75	0.98	0.90	0.80	0.79	0.64	0.77	0.77	0.60	0.41	0.67	0.50	0.44	0.32	0.47	0.38
Sint Maarten (Dutch part)					0.50											
Slovak Republic	0.76	0.85	0.78	0.74	0.50	0.61	0.48	0.13	0.39	0.40	0.47	0.26	0.45	0.47	0.45	0.29
Slovenia	0.69	0.43	0.67	0.52	0.64	0.12	0.46	0.85	0.70	0.76	0.70	0.29	0.96	0.89	0.76	0.58
Solomon Islands				0.61	0.04	0.60	0.74	0.30	0.77	0.60	0.34	0.37	0.50	0.39	0.62	0.35
Somalia				0.76	0.34	0.73	0.55	0.70	0.71	0.58	0.77		0.69		0.72	
South Africa	0.72	0.72	0.63	0.65	0.83	0.57	0.47	0.56	0.34	0.87	0.70	0.69	0.60	0.59	0.50	0.75
South Korea	0.38	0.74	0.26	0.59	0.89	0.58	0.69	0.65	0.85	0.84	0.91	0.68	0.66	0.26	0.43	0.29
South Sudan						0.54	0.00						0.85	0.04	0.85	0.06
South Sudan				0.61		0.38	0.43	0.77	0.62	0.75	0.61	0.43	0.49	0.59	0.39	0.77
South Sudan				0.66		0.25	0.26	0.39	0.65	0.48	0.42	0.46	0.36	0.42	0.31	0.44
Spain	0.80	0.75	0.62	0.47	0.72	0.65		0.60	0.74	0.22	0.63	0.17	0.72	0.57	0.72	0.54
Sri Lanka	0.93	0.90	0.93	0.51	0.91	0.56	0.36	0.93	0.73	0.58	0.72	0.25	0.50	0.58	0.45	0.77
Sri Lanka				0.76	0.72	0.70	0.34	0.30	0.63	0.44	0.54	0.51	0.47	0.57	0.49	0.51
Sri Lanka				0.62	0.64	0.55	0.52	0.37	0.87	0.49	0.66	0.62	0.57	0.48	0.57	0.42
St. Kitts and Nevis				0.84		0.71	0.71	0.67	0.59	0.42	0.50		0.38	0.41	0.38	0.23
St. Lucia	0.74	0.65	0.50	0.86	0.69	0.52	0.84	0.42	0.58	0.36	0.43		0.53	0.54	0.55	0.51
St. Lucia	0.11	0.17	0.61	0.65	0.70	0.57	0.65	0.45	0.44	0.80	0.72	0.33	0.38	0.64	0.33	0.64
St. Lucia				0.80		0.50	0.51	0.50	0.47	0.42	0.81	0.44	0.41	0.65	0.14	0.86
St. Lucia				0.44		0.59	0.49	0.59	0.21	0.29	0.63	0.52	0.63	0.64	0.56	0.63
St. Martin (French part)																
St. Vincent and the Grenadines				0.89		0.44	0.70	0.47	0.63	0.58	0.53		0.68	0.61	0.74	0.47
Sudan				0.49	0.66	0.38	0.57	0.55	0.71	0.34	0.28	0.60	0.66	0.57	0.57	0.59
Suriname				0.54	0.80	0.45	0.59	0.63	0.38	0.48	0.47	0.42	0.47	0.58	0.45	0.54
Swaziland				0.32	0.37	0.56	0.58	0.58	0.57	0.63	0.31	0.50	0.76	0.19	0.77	0.34
Sweden	0.95	0.64	0.92	0.65	0.80	0.63		0.53	0.62	0.61	0.78	0.45	0.63	0.43	0.39	0.33
Switzerland	0.65	0.78	0.89	0.50	0.79	0.58		0.61	0.68	0.26	0.85	0.67	0.33	0.33	0.33	0.57
Syrian Arab Republic				0.40	0.72	0.55	0.82	0.25	0.23	0.02	0.38	0.16	0.58		0.50	
Tajikistan				0.64	0.43	0.71	0.36	0.35	0.18	0.17	0.45	0.38	0.78	0.62	0.72	0.61
Tanzania				0.48	0.29	0.50	0.57	0.74	0.42	0.29	0.56	0.21	0.33	0.52	0.23	0.51
Thailand	0.53	0.80	0.60	0.89	0.79	0.77	0.41	0.54	0.85	0.77	0.55	0.78	0.67	0.73	0.39	0.32
Timor-Leste						0.33	0.56	0.40	0.14	0.39	0.42	0.58	0.07	0.48	0.14	0.57
Togo				0.33		0.66	0.53	0.55	0.55	0.44	0.44	0.70	0.50	0.72	0.62	0.64
Tonga				0.69	0.75	0.54	0.56	0.53	0.52	0.43	0.47		0.52	0.57	0.53	0.67
Trinidad and Tobago	0.33	0.28	0.64	0.45	0.80	0.69	0.62	0.64	0.50	0.30	0.70	0.78	0.69	0.38	0.62	0.56
Tunisia	0.59	0.59	0.50	0.69	0.66	0.59	0.63	0.28	0.43	0.41	0.39	0.27	0.44	0.20	0.35	0.47
Turkey	0.57	0.57	0.62	0.74	0.82	0.58	0.55	0.60	0.43	0.40	0.66	0.53	0.43	0.56	0.38	0.57
Turkmenistan				0.67	0.44	0.50	0.74	0.25	0.49	0.72	0.77	0.36	0.80	0.58	0.75	0.54
Turks and Caicos Islands							0.33									
Tuvalu				0.95			0.81	0.46	0.15	0.23	0.65		0.22	0.36	0.25	0.35
Uganda	0.10	0.50	0.10	0.76	0.95	0.44	0.68	0.75	0.62	0.47	0.76	0.24	0.69	0.35	0.48	0.47
Ukraine				0.58	0.56	0.67	0.25	0.43	0.56	0.67	0.62	0.42	0.44	0.27	0.40	0.26
Ukraine				0.56	0.70	0.54	0.51	0.44	0.65	0.34	0.73	0.49	0.49	0.59	0.47	0.51
United Arab Emirates		0.86	0.00	0.34	0.44	0.56	0.92	0.46	0.11	0.46	0.44	0.68	0.53	0.79	0.67	0.71

Country name	Labor force with primary education (% of total)	Labor force with secondary education (% of total)	Labor force with tertiary education (% of total)	Scientific and technical journal articles	Trademark applications, total	General government final consumption expenditure (% of GDP)	Net official development assistance and official aid received	Agriculture, value added (% of GDP)	Industry, value added (% of GDP)	Manufacturing, value added (% of GDP)	Services, etc., value added (% of GDP)	Unemployment, total (% of total labor force)	GDP growth (annual %) [NY.GDP.MKTP.KD.ZG]	GDP per capita, PPP (constant 2011 international $)	Previous GDP growth (annual %)	Previous GDP per capita, PPP (constant 2011 international $)
United Kingdom	0.85	0.66	0.49	0.68	0.67	0.43		0.57	0.47	0.49	0.76	0.55	0.37	0.84	0.41	0.47
United States	0.49	0.45	0.14	0.82	0.76	0.52		0.45	0.51	0.79	0.52	0.48	0.52	0.64	0.47	0.62
Uruguay	0.54	0.41	0.41	0.50	0.50	0.56	0.87	0.53	0.47	0.45	0.46	0.46	0.55	0.73	0.52	0.71
Uzbekistan				0.53	0.49	0.54	0.52	0.58	0.40	0.84	0.69	0.81	0.84	0.76	0.74	0.88
Vanuatu				0.56		0.48	0.55	0.29	0.42	0.46	0.39		0.55	0.28	0.64	0.28
Venezuela	0.27	0.81	0.00	0.20	0.61	0.41	0.78	0.44	0.33	0.71	0.58	0.49	0.67	0.42	0.68	0.35
Vietnam				0.68	0.64	0.65	0.65	0.66	0.65	0.53	0.89	0.61	0.64	0.66	0.69	0.44
Virgin Islands													0.30		0.34	
West Bank and Gaza	0.47	0.47	0.47			0.56	0.28	0.51	0.53	0.44	0.58	0.35	0.36	0.60	0.38	0.64
West Bank and Gaza				0.40	0.54	0.58	0.53	0.81	0.70	0.38	0.71	0.47	0.49	0.52	0.39	0.54
Yemen				0.61	0.30	0.33	0.62	0.40	0.56	0.53	0.48	0.47	0.97	0.43	0.94	0.47
Zambia				0.58	0.77	0.50	0.58	0.79	0.24	0.67	0.29	0.26	0.63	0.75	0.65	0.75
Zimbabwe				0.29	0.64	0.72	0.41	0.40	0.65	0.41	0.24	0.44	0.74	0.75	0.73	0.64

VITA

Fernando Javier Torre-Mora was awarded his BE in Computer Science from Universidad Simón Bolívar (USB) under the Aquiles Nazoa scholarship for academic excellence in 2013, and received a Fulbright scholarship to study his MS at the University of Missouri later that year. He spoke at the *IV Iberoamerican Congress on the Teaching on Engineering* and has, at the time of this publishing, submitted a paper to the *Journal of Information Sciences*. He additionally has two 2013 publications in the USB's arbitrated periodical *Universalia* and has experience with foreign language teaching, graphic design, software design, biomedical research, and library sciences.