Requirements Engineering Research: A Microcosm of International Economic Trends

Eustace Asanghanwa[1], Alan M. Davis[1], Oscar Dieste[1*], Ann M. Hickey[1], Natalia Juristo[2], Ana Moreno[2]

[1] University of Colorado at Colorado Springs, College of Business
PO Box 7150, Colorado Springs, CO 80933 USA
{enasangh,adavis,ahickey,odieste}@uccs.edu
[2] Universidad Politécnica de Madrid, Facultad de Informática
Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, SPAIN
{natalia,ammoreno}@fi.upm.es

Abstract. A major goal of requirements engineering (RE) research is to improve society's knowledge and practice of the RE discipline so as to positively impact society's ability to build new, ever more effective software systems. Since software systems now permeate almost every aspect of society, success in constructing such systems should benefit the economies of countries utilizing them. This paper begins to explore this new field of research, namely the macro-economic implications of RE research. Specifically, the present paper analyzes RE research publication trends (as a surrogate for RE research activity) by country and geographical region, and then compares those trends to key economic indicators. These analyses identified some very interesting correlations between publication trends and the economic indicators, encouraging us to explore the nature of these relationships in future research.

1 Introduction

Requirements engineering (RE) is the discipline of determining, analyzing, pruning, documenting, and validating the needs and requirements of stakeholders for a software system. RE research aims to improve society's knowledge and practice of the discipline so as to positively impact society's ability to build new, ever more effective software systems. While the relationship of RE and software system success has not been proven [1], it is certainly generally believed that more effective RE practices will increase the probability of success. For example, the Standish Group [2] showed that a large percentage of software systems that we build fail to meet the needs of the intended users. Boehm [3] and Jones [4] published data showing that a very large percent of project failures can be blamed on poorly understood requirements. Since the primary purpose of RE is to increase all parties' level of understanding of requirements, it seems reasonable to conclude that improving the practice of RE will increase the number of software system project successes. Moreover, since software systems now permeate almost every aspect of society,

* This author has been partially sponsored by the Spanish MECD/Fulbright Grant FU2004-0589.

success in constructing such systems should have a significant benefit to the economies of countries utilizing them. Consequently, if we believe that RE research positively impacts RE practice and therefore software system success, increases in RE research should have a positive economic impact.

This paper is part of an ongoing effort to investigate RE research and its impact on the global economy. RE publication rates are used as a surrogate for RE research activity. To determine the potential for this research, this paper analyzes RE research publication trends by country and geographic region, then compares major economic indicators for each geographic region to their RE publication rates to discover if correlations exist between economic and RE publication trends. If such correlations exist (as was found in this study), we will extend our research to explore the nature of the relationships between RE research and economic indicators. This paper has been structured as follows. In section 2, the methodology for this study is described. The preliminary results of the study are presented in section 3. Section 4 points out several issues for future research. Finally, section 5 summarizes what we have and have not learned from this research.

2 Research Approach and Limitations

The research was carried out using quantitative analysis of two sets of data: (1) a database of RE publications and (2) selected geo-economic data. The RE publications data was analyzed by country, geographic region and over time to identify publication trends. Correlation analyses were then used to identify potential relationships between RE publication rates and the geo-economic data. Several limitations existed because of either the quality of the data obtained or the analysis method used. These issues are discussed in the following sections.

2.1 Requirements Engineering Publication Data

The authors have collected bibliographic data on more than 2,500 publications in the area of requirements engineering over a period of 16 years. Data collection was performed using different methods as the compilation process evolved.

- Since 1989, Davis has been compiling an extensive bibliography of requirements-related publications. This bibliography was first published in his 1990 [5] and 1993 [6] requirements books.
- Since 1996, the bibliography has been available for public use on the web. It is now hosted at http:web.uccs.edu/adavis/reqbib.htm.
- Since 2001, the authors have been conducting regular online searches of IEEE Xplore, ACM Digital Library, ScienceDirect, Kluwer Online, Engineering Village 2, Wiley Interscience, and Springerverlag.com to search for papers whose abstracts or titles contained any combination of the words “requirement,” “specification,” “prototype,” “analysis,” “scenario,” and “use case.” The titles and abstracts of the resulting lists of publications were examined manually to determine relevance, and added to the bibliography as appropriate. References for approximately 75% of the

publications were also examined manually for RE publications not contained in the database. When new publications were detected, they were added to the database.
- Since 2004, the bibliography has been maintained (offline) in an Access database (referred to as "the database") for ease of analysis. Searches and reference lookups have been performed regularly up to the present day.

Currently, the database contains 2,645 unique publications spanning the years from 1963 through 2004 with a total of 5,551 authors recorded for those publications. All papers were reviewed manually to identify their authors' affiliations at the time of publication. We identified affiliations for 4,434, or 80% of the authors. Authors are located in 51 different countries. Each country was placed into a single large geographic region. We used six of the seven continents (excluding Antarctica) as regions, plus the Middle East which we isolated due to its current importance in international economics and politics.

2.2 Geo-Economic Data

The primary source of geo-economic data was the United Nations Environment Programme (UNEP). UNEP maintains an online database [7] of economic trends with data extracted from such prominent data providers as the International Labour Organization (ILO) [8], the International Telecommunications Union (ITU) [9], the Internet Software Consortium (ISC) [10], and the World Bank's World Development Indicators (WDI) [11], just to name a few. By country, the data collected included:

- Gross Domestic Product (GDP) (1960-2002).
- Gross Domestic Product per Capita (1960-2002).
- Industry Value Added (1972-1999).
- Service Value Added (1972-1999).
- Manufacturing Value Added (1986-1999).
- Agriculture Value Added (1972-1999).

The data was downloaded from the sources into Microsoft Excel where it was grouped into the same set of regions used for the RE publications.

2.3 Data Analysis

Microsoft Access queries were written to extract data from the RE publications database. Query results were then imported into Microsoft Excel to analyze publication trends and perform correlation analyses with the geo-economic data.

2.4 Limitations

The research approach used has three different types of limitations related to: (1) the completeness of the RE research data, (2) the accuracy and completeness of the geo-economic data and (3) the dependence on correlation analysis.

1. RE research data limitations:

Although the RE publications database contains an extremely large set of RE papers, some of the published research may have been omitted. There are several reasons why this can occur.

- Our database only includes publications in the English language. We know that many countries hold regional conferences in their national languages, and a few countries have software engineering journals published in their national languages. Therefore, countries where researchers commonly publish in non-English outlets are under-represented in the database.
- The online databases contain only limited historic data. For example, IEEE Xplore contains e-copies from IEEE Software, Computer, and Transactions on Software Engineering only as far back as 1988, COMPSAC back to 1978, HICSS back to 1988, IWSSD back to 1991, ICSE back to 1994, and APSEC back to 1997. This somewhat limited our ability to identify author affiliations for older papers, but had less impact on the general publication data since the data collection process included several alternatives for identifying publications.
- The decision on whether or not to include a candidate paper in our database was based primarily on reading the title and abstract. Thus, we may have erroneously included or excluded some papers based on incomplete knowledge of the actual contents of the papers.
- The field of RE is by no means well defined. It is thus possible we excluded publications which do not fit our viewpoint of what RE is.
- We continue to update the data. Thus, some totals or percentages may be not absolutely correct and could change slightly over time. However, the general trends and relationships have not changed by these updates.

Finally, we are using RE publications as a surrogate for RE research activity, but, not all research is published nor do all countries emphasize publication to the extent true for the United States. However, publication is a primary mechanism for disseminating research results to improve practice. And, it is through improvements in practice that we hypothesize RE research is able to impact the global economy. Therefore, we feel RE publications is a reasonable surrogate, at least for this preliminary analysis.

2. Geo-economic data limitations:
   - The cited geo-economic data is from a secondary source, the United Nations Environment Programme, and thus our results are only as accurate as that data.
   - The data from the United Nations Environment Programme is not always complete. For example, for some (mostly smaller or newer) countries, data is unavailable for either recent or distant years. However, smaller or newer countries tend not to be major generators of RE publications so any remaining errors represent little attenuation. More surprisingly, Value Added data was spotty for the U.S. and Europe. We have made every attempt to reduce the effects of these omissions in our analyses by seeking data from alternative sources, and have clearly indicated in our analyses when this was not successful.
3. Correlation analyses limitations:
   - The existence of a correlation between two variables does not mean that one variable influences or causes the other. For example, when we state that there exists a positive correlation between RE publications and some economic indicator (e.g., GDP), we do not mean that GDP is explained by the RE

publication trend (or vice versa). There are many possible reasons for this correlation including pure chance. We simply mean that some relationship does exist that may justify further research into the nature of that relationship.

– As mentioned above, we do not have complete data for several countries for some economic indicators. Correlation analyses were not run in those cases.

3 Results

3.1 National and Regional Demographic Trends

The fifty-one countries that represent the affiliations of the authors of the 2,645 publications are shown in Table 1. These countries are shown in decreasing order of total publication output. Although the USA is certainly responsible for a large number of total RE publications, a different perspective arises when we examine publications per million economically active capita (#/mc) as shown in the third column of Table 1. When looked at this way for countries with at least 5 publications, the UK and Sweden take the lead, with the USA, Ireland, and Austria close behind.

Table 1. RE publications by country

Country	#	#/mc	Country	#	#/mc	Country	#	#/mc
USA	1558	10.7	China	17	0	Latvia	2	1.5
UK	362	12.2	Ireland	16	10.0	Liechtenstein	2	*
Canada	127	7.7	Finland	15	5.8	Luxembourg	2	10.9
Germany	123	3.1	Switzerland	15	3.9	Macau	2	8.7
Japan	75	1.1	Taiwan	15	*	Malaysia	2	.2
Australia	69	7.1	Denmark	13	4.4	Morocco	2	.2
Spain	59	3.4	Portugal	12	2.4	Slovenia	2	2.0
Sweden	54	11.3	Greece	10	2.2	Turkey	2	.1
Brazil	49	.6	India	10	0	Venezuela	2	.2
Italy	42	1.7	Scotland	8	*	Chile	1	.2
France	39	1.5	Poland	6	.3	Croatia	1	.5
Austria	35	9.3	UAE	6	4.4	Iran	1	.0
Netherlands	34	4.6	Hong Kong	5	*	New Zealand	1	.5
Belgium	33	7.8	Argentina	4	.3	Pakistan	1	0
Israel	22	8.5	Singapore	4	2.0	Russia	1	0
Korea, South	19	.8	Cyprus	3	7.8	South Africa	1	.1
Norway	18	7.8	Costa Rica	2	1.2	Ukraine	1	0

* - Insufficient data

When we examine regional output, as shown in Fig. 1, we see that continental North America has made the most contributions, currently representing 59% of all publications in our database. Europe follows closely with 30%, and the least were made from Africa with less than 1% of the contributions. Notice that the totals of the figures in this chart add up to more than 100% because some publications had authors from multiple continents.

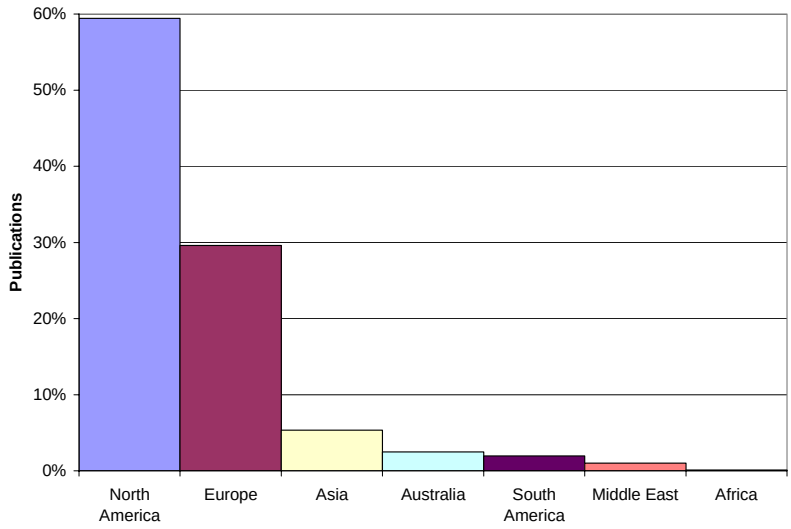

Fig. 1. RE publications per region

Fig. 2 presents data on regional growth patterns of RE publications by half-decade periods from 1965 through 2004 by region. From Fig. 2, it is quite easy to visually extrapolate RE publication output linearly for North America and Europe. In particular, it is clear that Europe is likely to surpass North America in annual RE publication production within just a few years.

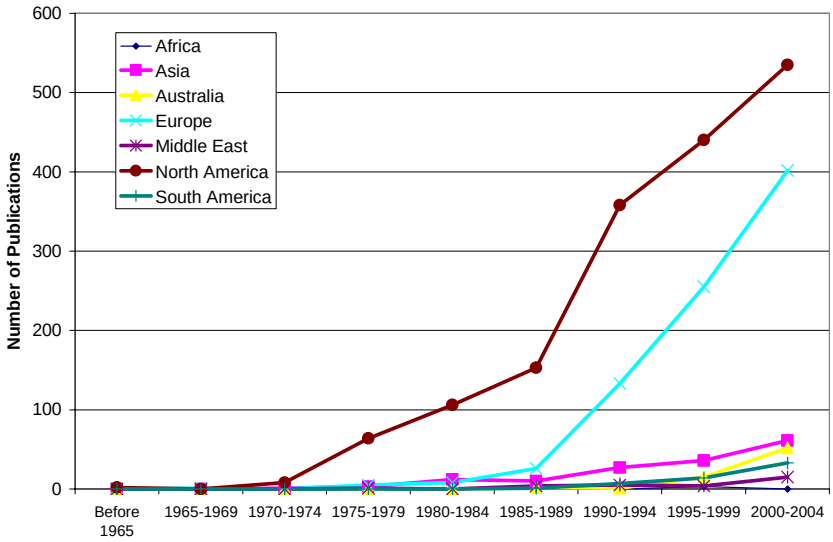

Fig. 2. Regional growth patterns of RE publications

3.2 Regional Geo-Economic Trends

Gross domestic product (GDP) is the total value of all goods and services produced within an area during a year. As such, it can be considered a general indicator of the relative economic health of a nation or region. Fig. 3 shows the average annual GDP by region. We ran a correlation between the data of Fig. 2 and Fig. 3 and discovered high positive correlations: Africa: .830, Asia: .725, Europe: .940, Middle East: .903, North America: .780, and South America: .657. This is logical considering our earlier assumptions that the role of RE research is to increase one's ability to perform RE, and performing RE well likely would lead to more effective software system developments and improved economic performance. This phenomenon is further substantiated by comparing the ranking of countries in Table 1 with the ranking of countries in order of decreasing GDP, as shown in Table 2. Data for the top 50 GDP producers and all 51 countries with RE publications are included in Table 2. Note just a few inconsistencies (i.e., ranking differences of 10 or more) when looking at the top 50 GDP producers: Argentina, China, India, Indonesia, Iran, Mexico, Russia, Saudi Arabia, South Africa, Thailand, and Turkey appear to under-produce relative to their GDP; whereas Ireland, Israel, Norway, Sweden, and UAE appear to over-produce relative to their GDP. On the surface, at least some of these inconsistencies seem related to the development status and economic activity level of these countries.

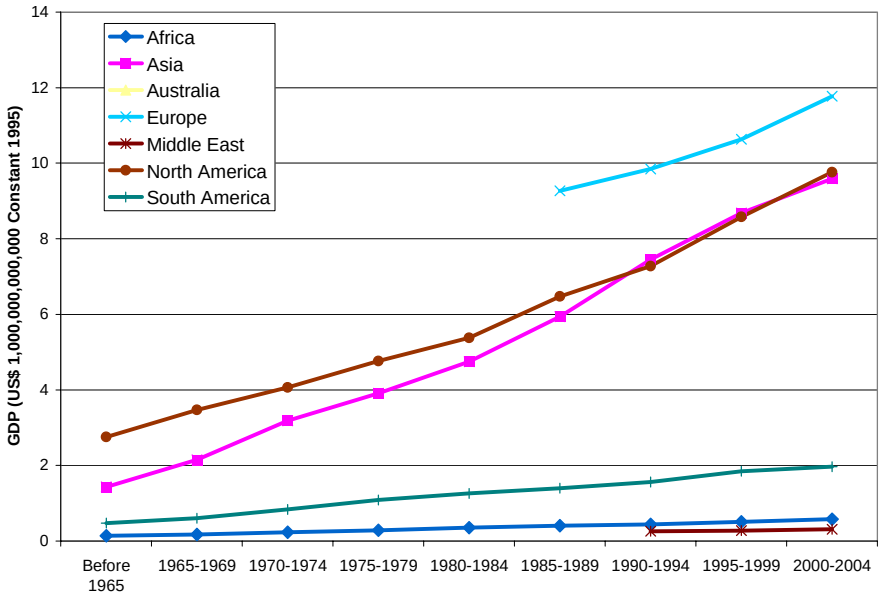

Fig. 3. Average annual GDP by region[1]

[1] GDP data not available for 2003-2004. We extrapolated 2001-2002 GDP data linearly.

Table 2. Countries ranked by GDP and RE publications

Country	GDP Rank	RE Rank
USA	1	1
Japan	2	5
Germany	3	4
France	4	11
UK	5	2
Italy	6	10
China	7	18
Brazil	8	9
Canada	9	3
Spain	10	7
Korea, S.	11	16
India	12	*25
Netherlands	13	13
Australia	14	6
Russia	15	*44
Mexico	16	n/a
Switzerland	17	*20
Belgium	18	14
Sweden	19	8
Austria	20	12

Country	GDP Rank	RE Rank
Argentina	21	*31
Indonesia	22	n/a
Denmark	23	23
Turkey	24	*34
Thailand	25	n/a
S. Africa	26	*44
Norway	27	17
Finland	28	*20
S. Arabia	29	n/a
Greece	30	*25
Poland	31	*28
Portugal	32	24
Ireland	33	19
Iran	34	*44
Malaysia	35	*34
Singapore	36	*31
Israel	37	15
Colombia	38	n/a
Philippines	39	n/a
Chile	40	*44

Country	GDP Rank	RE Rank
Egypt	41	n/a
Pakistan	42	*44
Venezuela	43	*34
N. Zealand	44	*44
Peru	45	n/a
Hungary	46	n/a
Czech	47	n/a
UAE	48	*28
Bangladesh	49	n/a
Algeria	50	n/a
Ukraine	51	*44
Morocco	52	*34
Luxembourg	58	58
Slovenia	61	*34
Croatia	62	*44
Costa Rica	72	*34
Cyprus	78	33
Macau	93	*34
Latvia	96	*34

*Ties

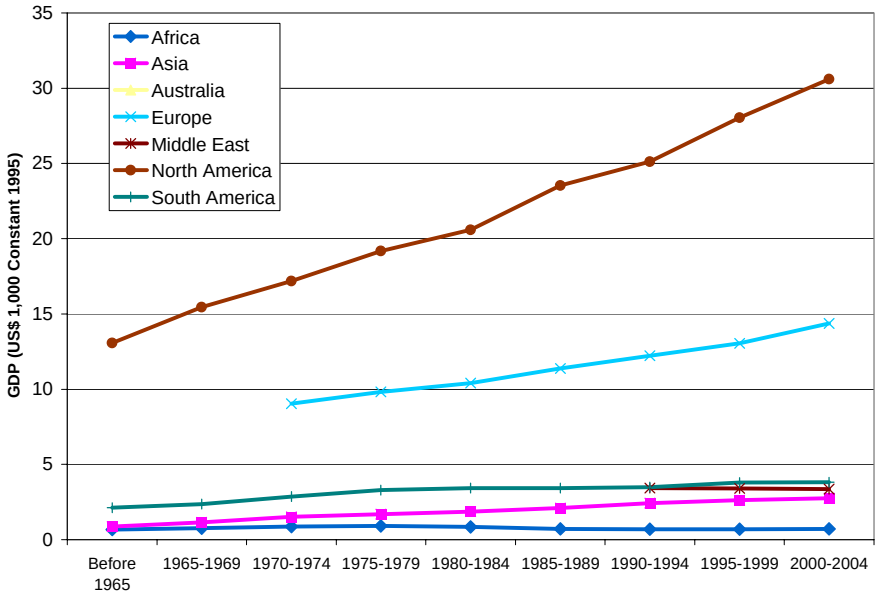

Fig. 4. GDP per capita by region

Gross Domestic Product per Capita normalizes GDP by the size of the economically active population. As such, it can be considered an even better indicator of the relative economic health of a nation or region. Fig. 4 shows the GDP per capita

by region. Once again, we ran a correlation between the data of Fig. 2 and Fig. 4 and found correlations which supported the proposed link to development status: Africa: -.417, Asia: .656, Europe: .816, Middle East: -.929, North America: .749, and South America: .507. Specifically, developed regions had strong positive correlations, regions with a mix of developed and developing countries had weaker positive correlations, and developing or under-developed regions had negative correlations.

Four economic values (Industry value added, Service value added, Manufacturing value added and Agriculture value added) indicate how active the industrial, service, manufacturing, and agriculture sectors are in the overall economy. Fig. 5 through Fig. 8 shows the patterns of sector activity by region. Complete data for these economic indicators was the most difficult to obtain, therefore results are spotty as indicated by the missing data on Fig. 5 through Fig. 8 and blank correlations in Table 3. However, even with the missing data, we felt it was important to take at least a preliminary look at these indicators to determine if additional data collection and analysis is warranted.

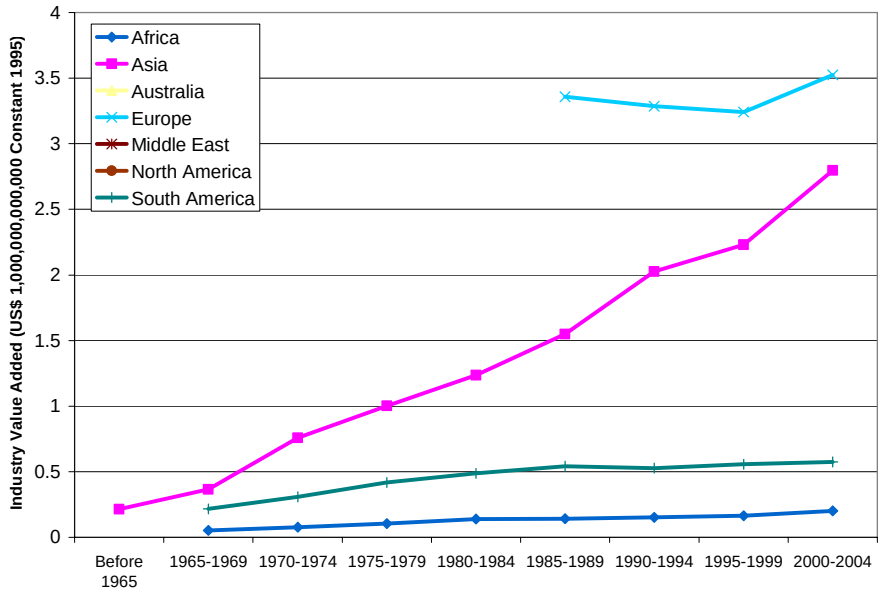

Fig. 5. Industry value added by region[2]

[2] Missing segments indicate missing data. Correlations were run only for periods with data.

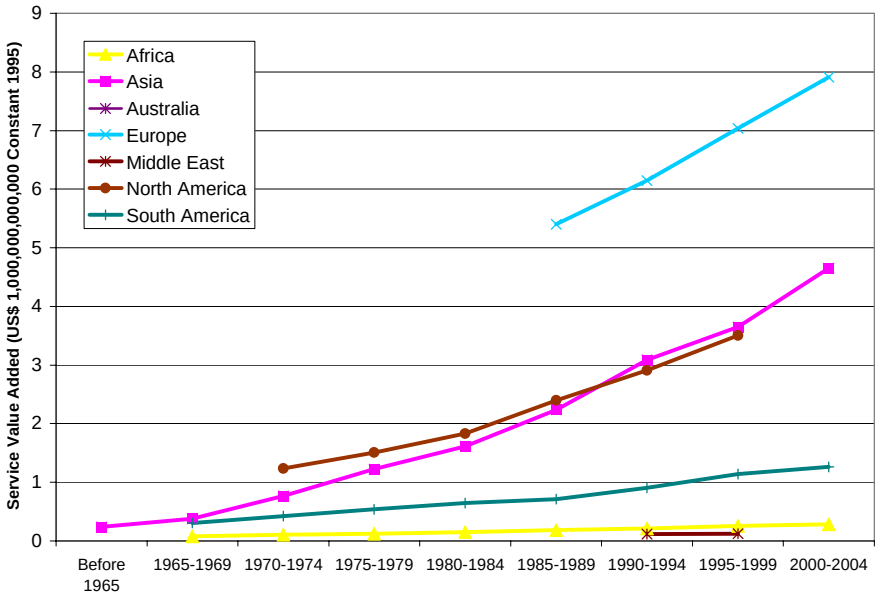

Fig. 6. Service value added by region[2]

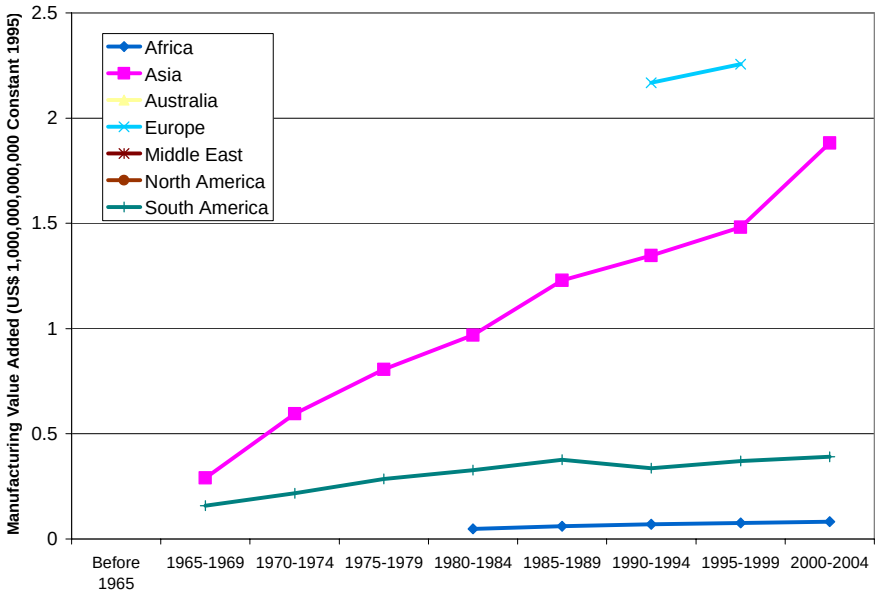

Fig. 7. Manufacturing value added by region[2]

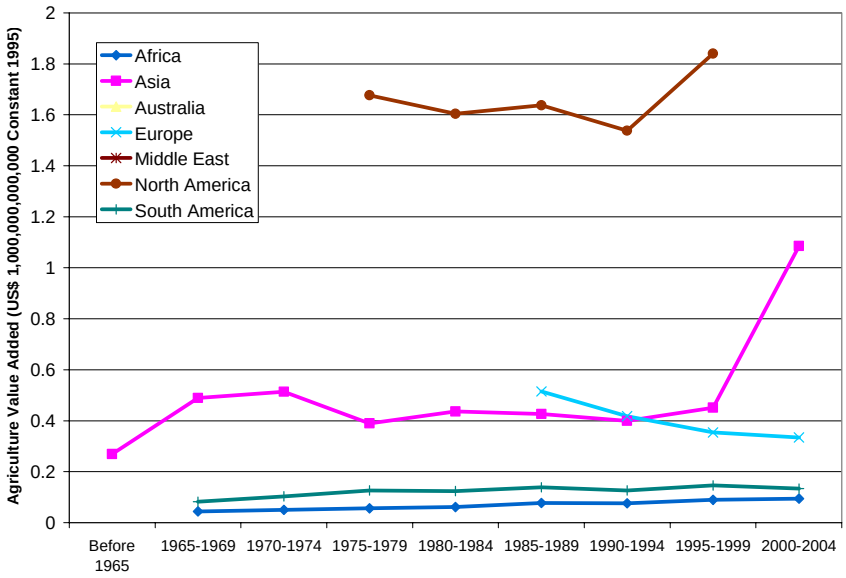

Fig. 8. Agriculture Value Added by Region[2]

Because requirements engineering is closely aligned with the service and industry sectors and somewhat with the manufacturing sector as it increasingly relies on automation, we should expect high correlations with the industry and service sectors and possibly the manufacturing sector, and relatively low correlations with the agriculture sector. Table 3 summarizes the correlations found by sector for each region where data was available. In general, we found high positive correlations with the industry, service, and manufacturing sectors. For agriculture, we found the expected negative correlation for Europe, low correlations for North and South America, but surprising high positive correlations for Africa and Asia. These and some of the other unexpected correlations (e.g., perfect 1.00 correlations) can be explained by a combination of low RE publication rates in Africa and the Middle East or a low number of data points for the economic indicators. Other anomalies will require further research to explain.

Table 3. Correlations between RE publications and several economic indicators

	Industry value added	**Service value added**	**Manufacturing value added**	**Agriculture value added**
Africa	0.789	0.888	0.922	0.844
Asia	0.830	0.795	0.763	0.934
Australia	-	-	-	-
Europe	0.831	0.893	1.000	(0.727)
Middle East	-	(1.000)	-	-
North America	-	0.963	-	0.255
South America	0.472	0.752	0.495	0.326

4 Future Research

There are several more issues that should be explored to gain a good understanding of the impact of RE in the economy. For example, we have performed and reported on correlations between RE output and six economic indicators listed at the beginning of this paper. Nevertheless, there are several other indicators that still need to be explored: Exports of Goods and Services; Imports of Goods and Services; Illiteracy Rate; Unemployment Rate (see following paragraph); Foreign Direct Investments; Internet Hosts; and Internet Users.

We are particularly interested in plotting unemployment levels by region. Unemployment levels are generally negatively correlated with economic prosperity. We should find a strong negative correlation between unemployment and RE publication output. However, a cynical assessor of RE research might be surprised by the negative correlation; after all, that individual might feel that RE research is conducted only when there is nothing else better to do, and thus would conclude that high unemployment would naturally lead to higher output of RE publications!

This same study could also be performed using RE-related patents as a surrogate for RE research activity instead of RE publications. Patents are another major output of research and may be even more linked to innovation than scientific publications. Thus it is foreseeable that, as predicted for RE publications, RE-related patents also contribute positively to the economic prosperity of nations.

Finally, this paper primarily analyzed major regional trends. We expect to perform similar analyses on individual countries in the near future. No doubt, specific countries are likely to (a) not be representative of their surrounding regions, or (b) have developed major expertise in RE caused, for example, by major national initiatives to excel in specific industries, e.g., India for software outsourcing. If needed, this detailed analysis can be expanded to concrete regions, e.g., Bangalore, India, where interesting.

5 Summary and Conclusions

This paper reports on an entirely new research area, namely the correlation of requirements engineering publication with worldwide economic and demographic trends. The results have shown strong correlations between RE publication output and certain economic indicators, providing preliminary support for a relationship between RE research and economic prosperity. However, it is too early to predict any causal relationship. The correlations could support the naïve hypothesis stated at the beginning of the paper, i.e., that increasing RE research may have a positive impact on the economic health in a region. Alternatively, they could also indicate that the healthier the economy of a region, the more likely it is to expend energy in the direction of publishing in the area of requirements engineering. Or, they may be completely explained by other indicators all together. However, the correlations are interesting enough to warrant continued research to further explore the exact nature of the relationships between RE research and global economic trends.

References

1. Davis, A., and D. Zowghi, “Good Requirements Practices Are Neither Necessary Nor Sufficient," to appear, Viewpoints, Requirements Engineering Journal, 2005.
2. Standish Group. The Chaos Report. www.standishgroup.com, 1995.
3. Boehm, B., et al., “Some Experiences with Automated Aids to the Design of Large-Scale Reliable Software,” IEEE Transactions on Software Engineering, 1, 1 (March 1975), pp. 125-133.
4. Jones, C., Applied Software Measurement, Second Edition, New York: McGraw Hill, 1991.
5. Davis, A., Software Requirements: Analysis and Specification, Prentice Hall, Englewood Cliffs, NJ, 1990.
6. Davis, A., Software Requirements: Objects, Functions, and States, Prentice Hall, Englewood Cliffs, NJ, 1993.
7. United Nations Environment Programme, www.unep.org.
8. International Labour Organization, www.ilo.org.
9. International Telecommunications Union, www.itu.org.
10. Internet Software Consortium, www.isc.org.
11. World Bank, World Development Indicators, www.worldbank.org/data/wdi2001.