Determinants of citation impact: A comparative analysis of the Global South versus the Global North

Maastricht University

UNITED NATIONS UNIVERSITY

UNU-MERIT

Working Paper Series

#2016-029

Determinants of citation impact: A comparative analysis of the Global South versus the Global North

Hugo Confraria, Manuel Mira Godinho and Lili Wang

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Maastricht Graduate School of Governance (MGSoG)
email: info-governance@maastrichtuniversity.nl | website: http://www.maastrichtuniversity.nl/governance

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology
UNU-MERIT

Maastricht Graduate School of Governance
MGSoG

Determinants of citation impact: A comparative analysis of the Global South versus the Global North

Hugo Confraria
UNU-MERIT, Maastricht University,
Boschstraat 24, 6211AX Maastricht, The Netherlands
confraria@merit.unu.edu

Manuel Mira Godinho[*#]
[*] ISEG (Lisbon School of Economics and Management), Universidade de Lisboa
Rua do Quelhas 6, 1200-781 Lisboa, Portugal
[#] UECE (Research Unit on Complexity and Economics)
Rua Miguel Lupi 20, 1249-078 Lisboa, Portugal
mgodinho@iseg.ulisboa.pt

Lili Wang
UNU-MERIT, Maastricht University,
Boschstraat 24, 6211AX Maastricht, The Netherlands
wang@merit.unu.edu

Abstract

The impact of the scientific output produced by different nations in different fields varies extensively. In this article, we apply bibliometric and econometric analysis to identify which countries are producing research with relatively higher scientific influence, and to understand what factors lead to higher citation impact. We focus specifically on the Global South because countries in this group are starting to converge in terms of output with the Global North. We find that previous citation impact, level of international collaboration and total publications in a specific scientific field are important determinants of citation impact among all nations. Yet, specialisation in particular scientific fields seems significantly more important in the Global South than in the Global North. We propose possible explanations for the patterns found and derive some policy implications.

Key words: Science, Global South, Development, Research policy, Bibliometrics, Scientific impact, Citation impact.

JEL codes: O38, O39, O57

1. Introduction

There is a widely held assumption that scientific research has positive effects on economic development, namely by increasing human capital, driving productivity growth or providing evidence to inform policies and practice (Salter & Martin, 2001; DFID, 2014). Yet, the process by which it happens is complex and there has been extensive debate about the extent to which development funders and governments in the Global South (or the peripheries) should invest in research.

A crucial aspect of analysing the scientific performance of countries is to understand if their scientific output is having international impact or influence. Studies that focus on measuring the scientific impact of countries usually use citation analysis mainly because it enables international comparisons to be more objective (Garfield, 1979). This can be regarded as one crucial aspect of scientific quality, and thus a "proxy" for quality as follows from the bibliometrics literature (Moed, 2005).

There are numerous studies in this field assessing research at the country level; however few try to understand what the determinants of citation impact are. This type of analysis can help to understand why some scientific systems are performing better than others, and this gap in the literature can be particularly relevant in order to create insights for science policy. Knowledge on these aspects will further the policy learning cycle and ultimately increase the accountability of public policies.

Using the *InCites*[TM] tool of Web of Science/Thomson Reuters (*WoS*[TM]), this article applies bibliometric and econometric analysis to evaluate which countries in the World are producing research with higher research citation impact, and to understand what factors lead to those higher results. The ability to estimate the expected number of citations of countries, taking into account country characteristics and other variables at the subject category level, can be particularly relevant for policy makers in low and middle-income countries (Global South), where public funds to finance the research system are scarce.

Our main objectives are, first to create a comprehensive framework that can be used in the interpretation of a country's citation impact, particularly in the Global South; second to contribute to citation theory by understanding if the citation impact indicators commonly used in high-income countries are adapted to be used in lower income contexts; and third to aid science policy-makers by identifying independent variables that influence significantly the citation impact of countries.

In what follows, we will first focus on the framework aspects of our analysis, then describe the data and methodology used, and afterwards we will discuss the results obtained. Finally, conclusions will be put forward.

2. Background

2.1. Science in the Global South

The North–South divide is generally considered on its political and socio-economic dimensions. Commonly, definitions of the Global North include North America, Western Europe and developed parts of East Asia, while the Global South is perceived has being made up of Africa, Latin America, and developing Asia including the Middle East. In this study we define Global North and Global South in two ways: firstly by using the World Bank definition of low & lower-middle-income countries versus upper-middle & high-income countries[1]; and secondly by dividing the World between OECD countries[2] and non-OECD countries. This possible division of the world into Global South and Global North has been perceived not only in terms of wealth or human development, but also in terms of scientific development.

In this context the understanding of the links between research investment and development has attracted an increasing attention. Though it has been recognized that there is no unique path to successful economic development that every country has to emulate, scholars like Lall (2000), Bernardes & Albuquerque (2003) and Fagerberg & Godinho (2004) have stated that in recent decades, countries that have caught up rapidly have tended to invest in their higher education system and have developed indigenous research efforts. According to Mazzoleni & Nelson (2007) the research programs that effectively contributed to catch-up did not operate within "ivory towers", but rather they were oriented towards an actual or potential user-community. They were projected to help solve problems, and advance technology, applicable to a particular economic area.

There are several ways which research carried out within national borders can help to provide both effective and focused responses to domestic problems, namely by being an enabler for providing up-to-date and qualified training for the new generations of university graduates, helping to attract qualified people to the country, and improving the quality of local advice to government and industry (Goldemberg, 1998). Investments in science can not only provide knowledge and skills for increasingly knowledge-intensive industries, but also generate a

[1] See the list of countries here: http://data.worldbank.org/about/country-and-lending-groups
[2] See the list of countries here: http://www.oecd.org/about/membersandpartners/list-oecd-member-countries.htm

"domestic base of good scientists, which can break into the international networks where new technologies are being hatched" (Nelson, 2005). These scientists can act as important conduits of frontier knowledge into the local academic research community (Barnard et al., 2012), which can potentially diffuse that knowledge to students, the economy and the general public.

Hence the "scientific culture" of nations (Godin & Gingras, 2000) has been recognised as a relevant dimension achieved through countries investing in science. As stated in the latest *UNESCO Science Report,* "the critical thinking that comes with science education is vital to train the mind to understand the world in which we live, make choices and solve problems. Science literacy supplies the basis for solutions to everyday problems, reducing the likelihood of misunderstandings by furthering a common understanding. It provides answers that are testable and reproducible and, thus, provides the basis for informed decision-making and effective impact assessments" (UNESCO, 2015).

These arguments reveal the importance of science for international development, though from an economic perspective one has to take into account the opportunity costs arising from investing in research. Therefore a necessary and integral part of science policy is to monitor and evaluate the various facets of the scientific enterprise. By measuring the different characteristics of the scientific systems, it is possible to create and manage policies to improve the scientific performance of countries.

2.2. Can the Global South use the same bibliometric indicators as those used in the North?

The use of bibliometric indicators for assessing the impact of scientific publications has been on the rise in recent years. The ability of it to lower costs and time of assessment without being invasive, enlighten political choices by doing international comparisons as well as their perceived objectivity, have been some of the main forces behind its growing popularity (Moed, 2005). However, the bibliometric assessment of research performance is based on a central assumption: scientists who have to communicate something important do publish their findings in international peer-reviewed journals. This choice introduces unavoidably a limited view of a complex reality (van Raan, 2004). For instance, journal publications are not in all fields the main carrier of scientific knowledge; there is language bias since most journals in *WoS*^{TM} are written in English; and countries have different levels of access to some journals, due to their financial constraints, selectivity or publication policies (Lawrence, 2003). This last limitation is particularly relevant in the Global South and may have acted in the past as a stimulus for researchers from those countries to seek publication through other channels, namely through other means that are not registered in *WoS*™, or in other similar databases. This problem was challenged recently by the Research4life[3] partnership, which intends to provide developing countries with easy access to peer-reviewed content. This initiative, which aims to reduce the "e-gap" between rich and poor countries, could contribute to a "normalisation" of access to the international circuit in the future. Yet this is still a limitation that we have to keep in mind when interpreting our results.

At the same time, both *WoS*™ and other indexing systems have enlarged considerably the database's coverage of Latin American and Caribbean (LA-C) journals in recent years. According to Collazo-Reyes (2014) the number of LA–C indexed journals in *WoS*™ has increased from 69 to 248 titles in just a period of four years (2006–2009). This unprecedented growth is related to a change in the editorial policy of *WoS*™. One example is the incorporation of *SciELO Citation Index*, wish includes regional journals from LA-C as well as titles from Spain, Portugal and South Africa, in their database.

For these reasons, and despite some recognized limitations, the use of bibliometric data and indicators has also been rising in the context of the Global South, where this type of analysis can be particularly relevant to understand successful processes of closing the S&T gap with the most advanced economies (Albuquerque, 2004).

2.3. Research citation impact

In line with this framework, one way to assess scientific impact is citation analysis. According to the seminal work of Garfield (1979) citations are the formal linkages between scientific publications that have particular points in common. They serve both instrumental and symbolic functions in the transmission and enlargement of knowledge (Merton, 1968). Instrumentally it tells us of work we may have not known before, some of which may hold further interest for us. Symbolically, it registers the intellectual property of the acknowledged source by providing a pellet of peer recognition of the knowledge claim, accepted or expressly rejected, that was made in that source.

According to Bornmann & Daniel (2008), two competing theories of citing behaviour have been developed in the past decades. One is often denoted as normative theory, and the other as the social constructivist view of citing behaviour. The normative theory, following Merton (1973) sociological theory of science, states that when a scientist cites a given article, he or she indicates

[3] http://www.research4life.org/

that the article was somehow relevant to the research performed. The citing author calls attention to some useful information included in an article, a method, a statistic, a result or other[4] and acknowledges intellectual or cognitive influence. Therefore, when a comparable article is cited more times than others it is considered to have more international scientific influence or impact (Moed, 2005).

The social constructivist view of citing behaviour challenges these assumptions. Constructivists argue that scientific knowledge is socially constructed through the manipulation of resources and the use of rhetorical devices (see Latour & Woolgar, 1979; Knorr-Cetina, 1981). In this view, citations are one rhetorical device that scientists employ to provide support for their papers and convince readers of the validity of their claims (Gilbert, 1977; Latour, 1987). According to the social constructivist perspective, citations perpetuate and shape existing patterns of institutional stratification and are little more than appeals to existing authority on the part of authors who wish to buttress their arguments.

To examine the legitimacy of these two theoretical approaches, some empirical tests have been undertaken. For example, by documenting significant positive effects of the cognitive relationship between citing and cited articles in terms of their common theoretical and topical content, as well as positive effects of a cited article's perceived quality and use of recent knowledge, Baldi (1998) provides support for the normative hypothesis. More recent empirical studies such as White (2004) and Riviera (2015) looking at the same problematic in different contexts, also give more importance to the normative point of view. These results give support to the idea that the behaviour of scientists is regulated by norms, which make the detection of citation patterns useful for the interpretation of bibliometric measures.

In our study, we assume that the normative theoretical approach is adequate to evaluate which countries in the World are producing research with higher scientific influence, and to understand which factors lead to those higher results.

[4] Authors also do self-citations, cite peers based on personal networks, flattery (citations of editors and potential referees) and do "negative" citations (contradicting other author). Yet, it is reasonable to assume that most citations are "positive", that is to say a sign of the fact that the citing author finds something useful in the material he cites. Deviating citation patterns, such as negative citations, can affect an analysis of an individual article or author, but this adverse effect tends to disappear in an analysis of larger aggregations of authors, such as departments, universities or countries (Moed, 2005).

2.4. Factors associated with higher levels of citation impact at a country level

In the literature, there are numerous studies assessing research at the individual, institutional and country level. Many other studies create and discuss new methods and metrics to evaluate citation impact. However, few try to understand what the determinants of citation impact are. Table 1 in appendix summarizes some of the factors that are known to be associated with higher citation rates at the article, author, institutional and country level. In our study we focus on the factors that are known to be associated with higher citation impact at the country level, namely: level of international collaboration (Narin et al., 1991; Glänzel et al., 1995; Katz & Hicks, 1997; Van Raan, 1998; Puuska et al., 2013), Wealth intensity[5] (King, 2004) and having English as an official language (Leimu & Koricheva, 2005). These determinants centre on ad hoc considerations and the literature has not, to the best of our knowledge, presented a comprehensive framework that could be used in interpreting a country's citation impact, particularly in the Global South. By bringing together the main arguments in this literature, this study aims to fill such gap in the literature.

In our analysis we will also include, as explanatory variables, previous citation impact, logarithmic scientific output, the percentage of publications in collaboration with industry and we will control for population size. Our argument regarding previous citation performance is that there might be path dependency, or the "Matthew Effect" (Merton, 1968) in science. Researchers, whose work has been highly cited in the past, are likely to receive more citations in the future. Regarding scientific output the rationale is that a higher scientific production, in the specific subject area, is a sign of higher critical mass that will foster quality and impact. This measure can also be used as a proxy for scientific specialisation since we are controlling for the total number of publications produced by a country. As for the percentage of publications in collaboration with the industry we intend to understand if citation impact is higher when the research performed by a country has a higher level of collaboration with the industry. This indicator can thus be seen as a measure of knowledge transfer between industry and academia, therefore if a country has a higher percentage of publications with at least one author from a corporation we assume that this country is performing more applied research[6]. At the same time,

---

[5] Gross Domestic Product per capita.

[6] An industry collaborative publication is one that lists its organisation type as "corporate" for one or more of the co-author's affiliations. However, not all single affiliations of all publications in *InCites*™ are unified as "university", "research institute", "corporate", etc. There are corporate affiliations that have not been unified yet not having an organisation type assigned and, therefore, are not identified as industrial collaborations. Large multinational corporations (MNE) have a higher probability of being identified and unified. Therefore, publications listed as industry collaborations are a lower boundary of the real co-publications activities. We would expect that countries

nations have obvious differences in size. To control for this we will also add as an independent variable logarithmic population.

3. Methodology

3.1. Data

Publication data were extracted from the *InCites*^{TM} (2014) platform provided by *Thomson Reuters*, which facilitates national comparisons across time periods. *InCites*^{TM} provides output and citation metrics from *WoS*^{TM} based on a dataset of more than 27 million papers from 1981 to 2014. The metrics for comparisons are created based on address criteria, using the whole-counting method, that is, counts are not weighted by number of authors or by addresses.

We adopted in our study the disciplinary break down of the *Essential Science Indicators* (ESI) areas. The ESI scheme incorporates a selection of journals made by Thomson Reuters. Our dataset covers 21 of the 22 ESI categories with a time span of 5 years (2008-2012). The research fields retained are as follows: Agricultural Sciences, Biology & Biochemistry Chemistry, Clinical Medicine, Computer Science, Economics & Business, Engineering, Environment/Ecology, Geosciences, Immunology, Materials Science, Mathematics, Microbiology, Molecular Biology & Genetics, Multidisciplinary, Neuroscience & Behaviour, Pharmacology & Toxicology, Physics, Plant & Animal Science, Psychiatry & Psychology, Social Sciences (general) and Space Science. The Multidisciplinary area was excluded since the publications included in this category could not be unambiguously classified into any of the 21 disciplinary areas.

The option for the ESI scheme took into account that there are several approaches to define a research field: on the basis of selected concepts (keywords), selected sets of journals, a database of field-specific publications, or any combination of these. The selection of a specific scheme[7] for the division of research fields has to take in account the trade-off between robustness of results and specificity of the subject category. Bibliometric data are characterised by skewed distributions, and hence robust statistics require considerable sample sizes. This favours using fewer categories, with more observations per category. However, articles in different subject categories have different citation propensities. Therefore the use of very broad categories (e.g. the 6 OECD categories scheme) can lead to differences in citation impact levels, which only reflect

with lower presence of MNEs have larger differences between the number of publications authored by the industry captured by *InCites*^{TM} and the real activity.

[7] InCites^{TM} provides six further schemes besides the 21 ESI based on a conglomerated of journals indexed in the WoS^{TM}, e.g. the 251 WoS^{TM} subject categories or the 6 OECD categories.

differences in the research portfolios of countries, as some countries are more specialized in fields within a given category that have a higher citation propensity.

We believe that the choice of the ESI scheme is the more adequate solution to solve this trade-off in this study. A common, although arbitrary, threshold is often a minimum of 50 full count publications for citation analysis. We use this threshold at a country/category level and we only consider countries that have at least 400 publications between 2008 and 2012.

A common debate in bibliometric studies is the use of social sciences and humanities for analysis (e.g. Marx & Bornmann, 2014; Hicks et al., 2015). The usefulness of citation impact indicators depends on the extent to which the research outputs are covered in bibliometric databases, and this coverage varies by subject category. The coverage tends to be higher in the natural sciences, which place a high priority on journal publications. In the social sciences and humanities, where the publication of books, book chapters, monographs, etc. is more traditional, the extent of the coverage is reduced. The 21 ESI categories include three categories related to social sciences, namely Economics & Business, Psychiatry & Psychology and Social Sciences (general), and exclude Humanities. Although the exclusive use of *WoS*™ data might not be appropriate to analyse citation impact in the social sciences, we decided that coverage was sufficient enough to include those three categories in our broad, country-level, analyses.

3.2. Approach and metrics

It is well known that different subject areas have different output propensities and that the publications belonging to each field have singular characteristics. Therefore, to be able to explain the different citation performances among countries and subject areas, we compute a multivariate regression analyses (OLS) with fixed effects at the subject area level.

Ordinary regression, assumes that all observations are independent. Yet, in our case, each country has 21 subject areas[8]. Because these potential 21 observations will share specific country characteristics, our observations are not independent from each other, and this would lead potentially to correlation of errors within countries, implying that the findings of statistical significance would be spurious. To tackle this we had to relax the independence assumption by clustering the errors at the country level (Moulton, 1990; McCaffrey et al., 2012).

[8] Because we use the threshold of at least 50 publications per subject areas, not all countries have 21 observations. This may lead to selection bias since in the countries that don't fulfil the threshold for the 21 observations; the categories that are being computed are potentially the ones that the country performs better. Nevertheless, this is just a hypothesis that needs more research to be fully understood.

When interpreting the results presented in this study it should be kept in mind that indicators measuring citation impact capture the influence of journal articles in the scholarly communication system. As a consequence of the partial and one-dimensional nature of these impact indicators, it is recommended to use more than one single indicator to obtain more robust conclusions (Bornmann & Leydesdorff, 2013). Consequently, for this study our dependent variable will be measured by two different indicators: (1) the share of highly cited publications, which shows the proportion of publications belonging to the top ten percent most cited documents in a given subject category, year and publication type. (PPtop10%), and (2) the field normalized citation score, which calculates the mean citation rate to a country's set of publications in a specific subject area, period of time and document type, divided by the mean citation rate of all publications in that subject area/period/document type (CXC). Both these variables are normally distributed indexes with some outliers in the right tale:

$$PPtop10\%\ (\%) = \frac{PPtop10\%\ (n)}{P} \qquad (1)$$

$$CXC = \frac{\sum_{i=1}^{P} c_i}{\sum_{i=1}^{P} [\mu_f]_i} \qquad (2)$$

Currently, there are several ways to calculate citation rates; from basic calculations like raw citation counts, citations per publication or the h-index, to normalized methods controlled for research field, publication year and document type as the "crown indicator", field normalized citation score (Waltman et al., 2011), percentile-based approaches (Pudovkin & Garfield, 2009), source normalized indicators (Waltman & Eck, 2012), among others. Since publications belonging to different subject areas have different propensities for being cited (Peters & van Raan, 1994; Bornmann et al., 2012), we use normalized indicators. The use of percentile ranking and field normalized citation score can avoid bias toward large size of country or field. Both this indicators can be computed by using the consistent *InCites*^{TM}/*Thomson Reuters* databases, to which we had access.

In our model (3), I is a measure of citation impact in a certain period *t*, subject area *s* and country *c*. I is a lag dependent variable in the previous period, *O* is the number of articles and reviews in *WoS*^{TM}, *IC* is the percentage of publications of a country in international collaboration and *IND* is the percentage of publications of a country in collaboration with industry. *C* is a set of country controls including total output, gross domestic product per capita (GDPpc), population size and English as an official language. Finally, α is the constant and ε is the unobserved residual.

$$I_{cst} = \alpha + \mu I_{cst-1} + \lambda O_{cst} + \eta IC_{cst} + \varphi IND_{cst} + \beta C_{ct} + \varepsilon_{ct} \qquad (3)$$

Some of our independent variables have an exponential distribution (GDPpc, population size and number of articles). We decided to apply logarithms in those cases. As for the multicollinearity problem, none of our independent variables is highly correlated with another (>60%).

Variables such as R&D intensity or numbers of researchers were dismissed as typically they are highly correlated with GDPpc or the number of articles. Furthermore, for many countries the availability or reliability of this type of data is dubious. In an earlier phase we also included in our model the variable "percentage of individuals using the internet", provided by the *International Telecommunications Union,* as a proxy for level of access to scientific journals. Yet, this indicator is also highly correlated with GDPpc.

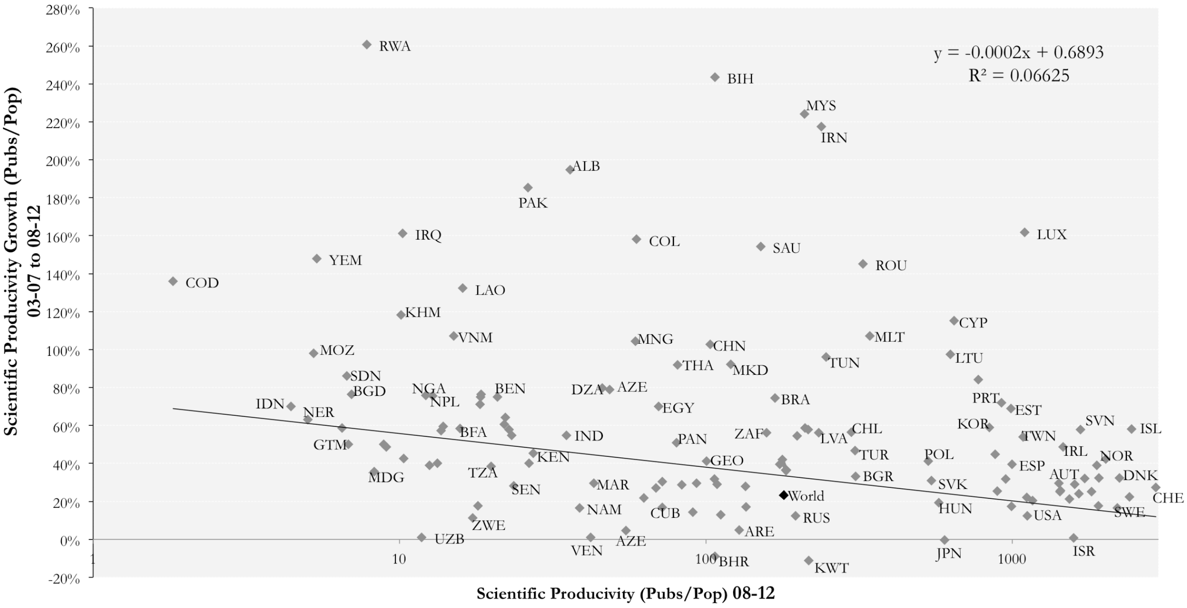
4. Results

4.1. Global Trends

World's long-term publication output in *WoS*[TM] has increased at an average rate of 3.5% since 1981. This growth rate has increased in the decade between 2004 and 2013 to an average of 5%. In 2013 the EU28 was still the world leader for publications (35%), followed by the US (27%), China (15%) and Japan (6%). Despite these impressive figures, the world shares of the EU28, US and Japan have fallen over the preceding decade. This decline was not due to the reduction of their scientific productivity (number of publications per population size) but due to the higher growth rates of other raising players as China or Brazil.

In Fig.1, by showing the scientific productivity growth rates of 132 countries between 2003-2007 and 2008-2012 versus their scientific productivity in 2008-2012, a modest trend of convergence denoted by the negative slope of the adjusted line is observable.

Fig. 1 Growth rate versus level of scientific productivity (publications per population).



Source: Own calculations based on *InCites*^{TM}.
Note: Vertical axis shows the growth rate of Pubs/Pop in 2008-2012 (matched with 2003-2007 level). Horizontal axis shows number of publications per million people (Pubs/Pop) in 2008-2012 (yearly average).

The Chinese case is impressive. China's scientific publications have more than doubled over the past ten years and, if following this trend, it will become the top producer of scientific publications in the world in the next ten years. This rapid growth reflects the coming of age of the Chinese research system, be it in terms of publications, number of researchers or investment (UNESCO, 2010).

As for Brazil, its share of world scientific output has increased at a constant rate from 1993 to 2006, followed by a fast rise in 2007 and 2008 to the level that Brazil shows in 2013. Vargas et al. (2014) argues that, in areas such as Agricultural Sciences Brazil's output increase since 2006 was mainly due to the expansion of Brazilian journals in *WoS*^{TM} and an increase in the number of issues published by these journals. This phenomenon may have led to more publications but fewer citations since journals edited in Portuguese have less international visibility.

Iran presents another remarkable story. This country more than tripled its number of publications between the two periods analysed. According to Akhondzadeh (2013), "scientific progress in Iran over the past few years was the result of the country's recent policies and programs to develop knowledge and facilitate researchers' access to the world's top academic resources".

In general the world figures show a global converging trend in science in quantity of publications. This result may be inflated by changes in size of the database, although we do not know the extent to which this may be the case. *WoS*^{TM} was significantly expanded between 2005 and 2010 in order to enlarge the regional coverage (Testa, 2011), and also in response to competition from *Scopus*^{TM}, which entered the market in 2004. Despite these relatively recent expansions of *WoS*^{TM} being possibly one reason behind the convergence that has been noted on scientific publication worldwide, a similar convergence trend has also been observed for R&D investment by the public sector between the Global North and Global South (UNESCO, 2015). Possibly, as Radosevic & Yoruk (2014) argue, these trends are associated with a change in scientific absorptive capacity of countries in the Global South. While 'absorptive capacity' has been generally defined as "the ability to learn and implement knowledge" (Cohen & Levinthal, 1990). Radosevic & Yoruk (2014) have defined absorptive capacity in the context of scientific research as "the ability to recognize the value of new, external information, assimilate it, and apply it in another context". In accordance to this view, researchers recombine and re-contextualize existing scientific knowledge and are able to generate novelty through their new publications.

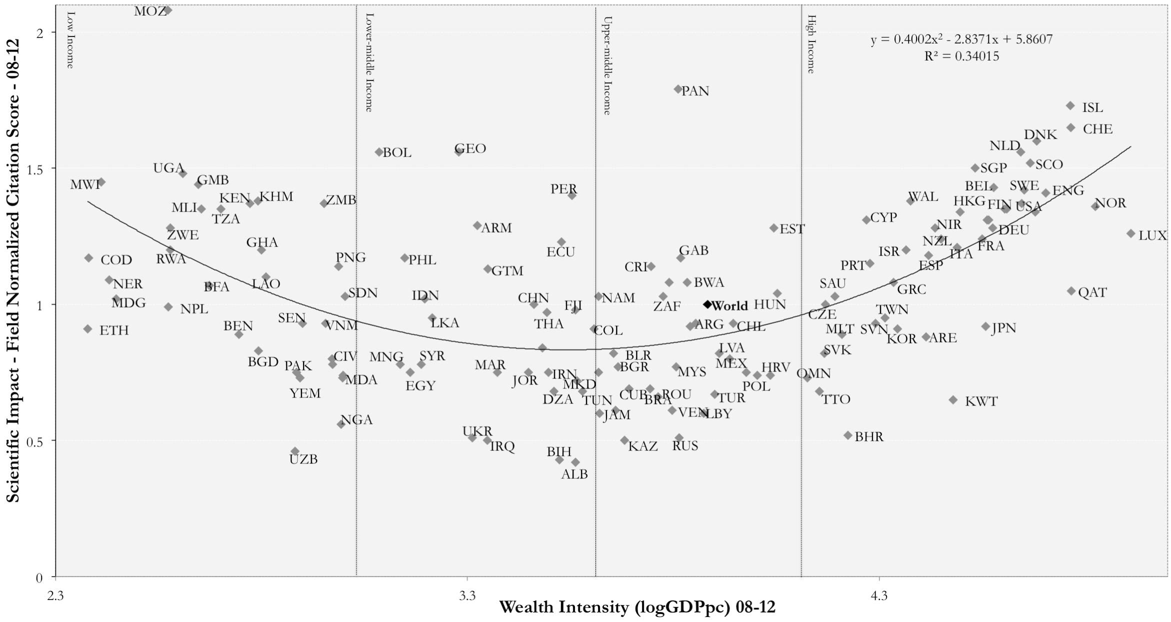
There is still a huge gap to overcome between higher-income and lower-income nations, yet the convergence which has been noted over the most recent years is certainly happening because some countries in the Global South are expanding their scientific capabilities and increasing their presence in scientific journals with high international visibility. Such changing trends provide some support to our quest to understand the determinants of citation impact in the Global South, despite that we analyse this by using indicators that are normally used to assess science in the Global North.

4.2. Wealth Intensity versus research citation impact

In this study we are particularly interested in understanding if at different levels of GDPpc there are different determinants of citation impact (scientific influence). To have a general overview of the relation between wealth intensity and citation impact we scatter in Fig. 2 and Fig. 3 the relation between average GDPpc and citation impact measured respectively by CXC and PPtop10% between 2008 and 2012, for countries that have more than 400 publications.

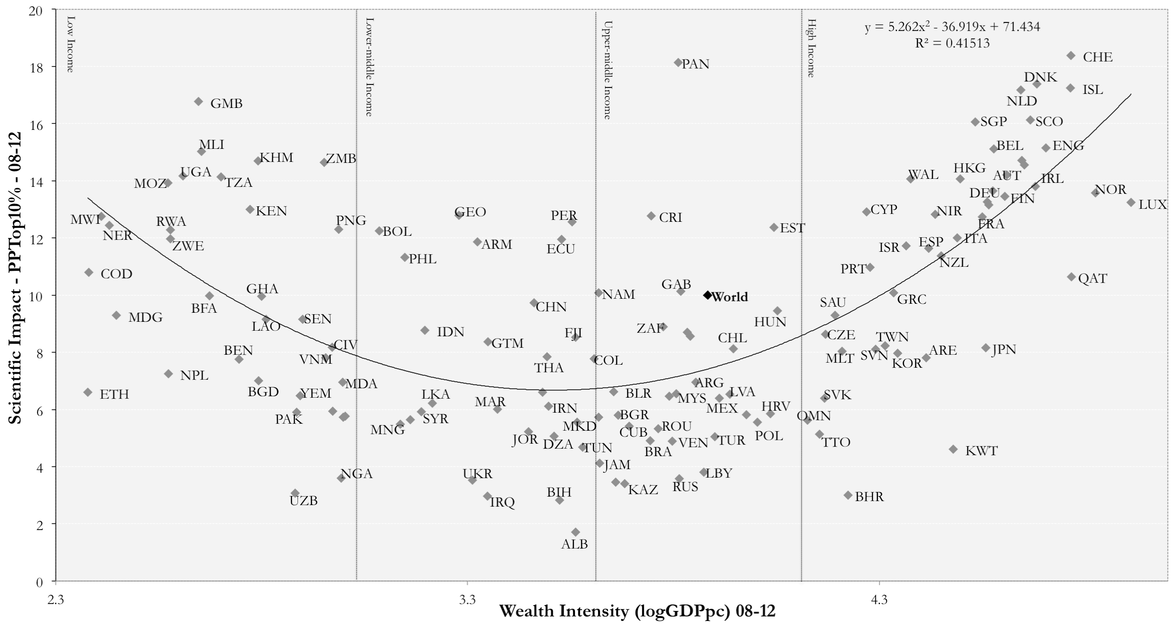
In both graphs we can observe a U-shaped pattern, with the adjusted lines having their inflexion points close to the *World Bank* borderline that divides "low & lower-middle-income" countries from "upper-middle & higher-income" countries. For "low & lower-middle-income" countries, the citation impact performance seems to follow a downward trend, though with substantial deviations from the curve. For "upper-middle & higher-income" countries, there seems to be a positive relation between the two variables. Such upward trend was already revealed by King (2004). In contrast, the U-shape pattern of our data seems to suggest that a nation's wealth only correlates positively with citation impact after a certain level of GDPpc.

Fig. 2 Citation Impact (CXC) versus GDPpc (2008-2012) [a]



Source: Own calculations based on *InCites*[TM]& World Bank
Note: Vertical axis shows citation impact in 2008-2012; Horizontal axis shows logarithm GDPpc (constant 2005$) in 2008-2012 (yearly average).

Fig. 3 Citation Impact (PPtop10%) versus GDPpc (2008-2012) [a]



Source: Own calculations based on *InCites*[TM] & World Bank
Note: Vertical axis shows citation impact in 2008-2012; Horizontal axis shows logarithm GDPpc (constant 2005$) in 2008-2012 (yearly average).

One should be aware that the countries showed in this graph have different dimensions. If a small country, in terms of publication output, has a set of publications that are very influential, the citations received by the articles produced by those researchers will improve significantly its citation intensity score. Mozambique, for example, can be one of those cases. Although its total production normalized by population is very low when compared to the world average (5 vs. 179 yearly publications per million people), its CXC is two times higher than the world average and the PP top 10% is close to 14%. In Mozambique, from 2008 to 2012, 95% of the country's publications have a foreign author. The high levels of citation impact in Mozambique may be happening because a small national scientific group is producing scientific publications with highly reputed international co-authors (Confraria & Godinho, 2014).

Another outlier in our graphs is Panama. There's also a story behind this case. Its citation impact (in intensity) is 79% higher than the world average for the CXC indicator, and 18% of its publications are in the top ten most highly cited papers. If we take a close look to the most highly cited publications from Panama between 2008 and 2012, we will find the most of them come from researchers affiliated to the *Smithsonian Tropical Research Institution.* This organisation is a bureau of the *Smithsonian Institution* based outside of the United States, which is dedicated to understanding biological diversity. According to their website[9] their "facilities provide a unique opportunity for long-term ecological studies in the tropics, and are used extensively by some 900 visiting scientists from academic and research institutions in the United States and around the world every year". In a country like Panama, which had a scientific output close to 1500 publications in the five years analysed, the presence of this research institute can make the difference. They function as a hub of attraction of world leading scientists, certainly having a huge influence on the high citation impact of Panama.

On the right edge of the U curve we find high-income countries such as Switzerland, Denmark, Iceland and The Netherlands. These are all relatively small European nations, which have been leading performers in this indicator for quite some time. For example, in one of the first studies analysing this issue, May (1997) also found that these countries were already leading the World in terms of "citation intensity".

In summary, our descriptive analysis suggests that higher levels of international collaboration may be extremely relevant for countries with both low GDPpc and smaller scientific communities. It is also perceptible that as middle-income countries may have more resources and larger scientific communities, they are not engaging so much on overseas collaboration. Our

[9] http://www.stri.si.edu/english/about_stri/index.php

U-shape trend indicates that this improvement in GDPpc, from low- to middle-income status, leads on average to lower citation impact. Finally, for high-income countries, both higher levels of GDPpc and small country size, which is again highly correlated with international collaboration intensity, seem to be critical factors.

4.3. Regressions analysis

We used *Stata*™ (StataCorp, 2013) to compute the multilevel regression (OLS) with fixed effects at the subject area level and errors clustered at the country level. The determinants of citation impact for publications between 2008 and 2012 were examined for 21 subject areas for countries with at least both 50 publications in a subject area and a total of 400 publications. After applying these restrictions, 126 countries and 1686 observations compose our global sample (see Table 2 for descriptive statistics).

Table 2. Descriptive Statistics

Variable	N	Mean	SD	Min	Max	Correlation									
						1	2	3	4	5	6	7	8	9	10
1) PPtop10%_0812	1686	9.39	5.28	0.00	55.51	1.00									
2) PPtop10%_0307	1686	8.45	4.73	0.00	32.81	**0.78**	1.00								
3) CXC_0812	1686	1.02	0.44	0.09	5.75	**0.90**	**0.69**	1.00							
4) CXC_0307	1686	0.92	0.34	0.03	3.09	**0.78**	**0.93**	**0.71**	1.00						
5) Pubs Area (log)	1686	3.03	0.72	1.72	5.57	**0.24**	**0.29**	**0.19**	**0.29**	1.00					
6) Internat. Collab	1686	56.07	19.55	9.66	100	**0.34**	**0.23**	**0.36**	**0.25**	**-0.51**	1.00				
7) Industry. Collab	1686	1.75	2.50	0.00	23.38	**0.35**	**0.28**	**0.32**	**0.29**	**0.13**	**0.09**	1.00			
8) Total Pubs (log)	1686	4.40	0.77	2.63	6.24	**0.20**	**0.26**	**0.13**	**0.26**	**0.87**	**-0.52**	**0.17**	1.00		
9) English Official	1686	0.25	0.43	0.00	1.00	**0.20**	**0.17**	**0.18**	**0.19**	**0.09**	0.03	-0.02	**0.08**	1.00	
10) GDPpc (log)	1686	3.95	0.62	2.38	4.91	**0.38**	**0.36**	**0.30**	**0.38**	**0.47**	**-0.21**	**0.31**	**0.55**	-0.02	1.00
11) Popul. (log)	1686	7.27	0.67	5.50	9.13	**-0.20**	**-0.12**	**-0.20**	**-0.14**	**0.42**	**-0.27**	**-0.12**	**0.46**	**0.07**	**-0.39**

Note 1: Correlation with bold numbers significant at p<.05;
Note 2: The numbers 0307 and 0812 in the variables stand for the time periods 2003-2007 and 2008-2012.

Previously in Fig. 2 and Fig. 3 we have seen that for different levels of GDPpc (below and above world average level), there are different patterns of citation impact. To understand if these differences are substantive, for the purpose of our analysis we split our sample in two groups of countries, following the World Bank's definition of "low & lower middle income" countries (Global South) and "upper middle & higher income" countries (Global North). Further we also introduce another North-South distinction, namely being or not an OECD country. Regressions were carried out for each of these groups separately. Generally the results for both North and South specifications are robust.

Table 3 reports the effect of the predictor variables on citation outcomes, using two dependent variables, respectively PPtop10% and CXC, for the citation rates. The South samples include 54 "low & lower-middle-income" countries (490 observations) and 89 non-OECD countries (928 observations), while the North samples include 72 "upper middle & higher income" countries (1196 observations) and 37 OECD countries (758 observations)[10]. Our model not only identifies variables that are significant in predicting higher levels of citations, but also identifies the relative contribution of each independent variable to the citation rates of countries.

Table 3. Determinants of citation impact in the Global South and the Global North

	Dependent variables							
	PPtop10%_0812				**CXC_0812**			
Independ. Variables	South (GDPpc)	North (GDPpc)	non-OECD	OECD	South (GDPpc)	North (GDPpc)	non-OECD	OECD
PPtop10% _0307	0.581*** (0.061)	0.583*** (0.046)	0.557*** (0.048)	0.692*** (0.053)				
CXC_0307					0.643*** (0.107)	0.561*** (0.067)	0.576*** (0.076)	0.687*** (0.086)
Pubs Area (log)	2.281*** (0.572)	1.189*** (0.335)	2.049*** (0.404)	0.944** (0.402)	0.156** (0.062)	0.015 (0.035)	0.128*** (0.040)	-0.013 (0.043)
Int. Collab	0.082*** (0.013)	0.090*** (0.011)	0.091*** (0.010)	0.087*** (0.010)	0.007*** (0.002)	0.007*** (0.001)	0.008*** (0.001)	0.007*** (0.001)
Ind. Collab	0.317*** (0.106)	0.186** (0.090)	0.268*** (0.095)	0.188* (0.107)	0.053** (0.023)	0.012** (0.006)	0.031*** (0.011)	0.010 (0.006)
Total Pubs (log)	0.144 (0.414)	1.211*** (0.423)	0.557* (0.315)	0.253 (0.758)	0.047 (0.054)	0.209*** (0.049)	0.104*** (0.036)	0.131 (0.08)
English Official	1.267*** (0.445)	0.651** (0.271)	0.929*** (0.324)	1.001*** (0.338)	0.059 (0.051)	0.043** (0.020)	0.050* (0.029)	0.066*** (0.021)
GDPpc (log)	-0.171 (0.592)	0.634 (0.435)	-0.067 (0.352)	0.146 (0.650)	-0.097 (0.063)	-0.002 (0.036)	-0.075** (0.033)	-0.034 (0.053)
Popul. (log)	-1.053** (0.410)	-1.723*** (0.345)	-1.292*** (0.336)	-0.820 (0.537)	-0.099** (0.041)	-0.176*** (0.030)	-0.122*** (0.031)	-0.099** (0.048)
Constant	-0.277 (2.652)		-0.165 (2.853)		0.342 (0.220)		0.302 (0.239)	
Observs.	1,686		1,686		1,686		1,686	
R-squared	0.730		0.729		0.649		0.644	

Note 1: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
Note 2: Linear regression, absorbing indicators. Std. error adjusted for 126 clusters (countries).

These results show that, in both groups of countries, previous citation impact, level of international collaboration and number of publications in the specific area are strongly associated with higher citation rates.

The first of these results indicates that despite the fast growth of some countries in recent years, globally a strong path-dependency in citation impact still holds. This may happen for specific reasons. It is well known in the literature that better known scientists tend to receive more credit

[10] We count England, Scotland, Wales and Northern Ireland as separate nations.

than less well-known scientists, even if their work is similar (e.g. Merton, 1968). Frequently cited researchers generally have higher status than researchers who are cited less frequently. Because status influences perceptions of quality, those with high reputation can increase their odds of keeping being cited, thus reinforcing further their status. If we admit similar self-reinforcing mechanisms exist at a more aggregate level, we can argue that nations performing better are also more likely to attract more tangible resources, such as research funding and outstanding graduate students, which can result in research of better quality and perpetuation of higher levels of citation impact. Our models do not suggest big differences between South and North in this regard, however it shows that previous performance is strongly associated to future citation impact; for example, for the PPtop10% indicator, countries that have 1% more papers in the top 10% more cited publications than others, have around more 0.6% papers in that same "excellent" tier in the next period.

As for the scientific output variable the countries that produce more publications in specific subject areas, have also higher citation rates per paper. This is intuitive since, in theory, these are the subject areas in which countries have a higher scientific capacity. At the same time, given the scale effect that arises, researchers in the same subject areas probably cite more frequently their own compatriots increasing the number of citations received by their country. This covariate represents not only the scientific output, but also the intensity of involvement in scientific activities of a country in a specific area (as gross expenditure on R&D and number of researchers are usually highly correlated with number of publications). Since this effect is significantly higher in the Global South, one implication of this result is that the importance of generating a higher critical mass in a specific field, in order to produce research with more influence in the World, seems to be larger in the South. For instance, countries in the South that have 50% more publications in a subject area than others, have on average 1% more papers in the top 10% most cited publications in the world. In the North this relation is significantly smaller.

Concerning international collaboration, it is well known in the literature that citation impact is typically greater when research groups collaborate, and the benefit strengthens when co-authorship is international (Van Raan, 1998). The rationale behind it is that scientists are likely to develop new and alternative ways of thinking when they interact with other scientists with diverse areas of expertise and backgrounds (Hollingsworth, 2006). Co-publication allows access to a larger social network that consequently leads to increased visibility which in turn is reflected in higher citation rates (Goldfinch et al., 2003). This cross-fertilisation is amplified by international collaboration because scientists who produce co-authored papers with foreign scientists are more likely to belong to elite research groups within their own countries (Adams, 2013). Since countries in the Global South depend a lot on international scientific networks in order to produce research that has visibility and impact (see Fig. 4 in appendix), we would expect that this positive relation is higher in the countries from the Global South. Yet our results seem to show that the importance of international collaboration is not significantly different in both groups of countries. Specifically, countries that have 10% more publications co-authored internationally in a subject area have, on average, 0.85% more publications in the top 10% in that subject area.

An interesting finding is that industry collaboration seems to matter for citation impact, especially in the Global South. Because most co-publications with industry are co-authored by staff at the large R&D-intensive technology companies in science-based industrial sectors such as biotechnology, pharmaceuticals, electronics, chemicals, and computers (Godin, 1996), this indicator can be seen as a "knowledge linkage indicator" (Tijssen, 2012) between multinational R&D-intensive technology companies and public research organisations. This type of collaboration with industry is very likely to be driven by the need for access to international R&D networks, advanced research facilities, and contributions by scientists and research teams of international repute. Whereas from the industry side, researchers may be attempted to publish because they aspire to be active members of a research community and want to be regarded as such by their peers, alongside other objectives to make corporate research findings public (Godin, 1996; Tijssen et al., 2009). The industry side may feel a particular appeal to collaborate with scientists from the South as a way to reach specific resources or for testing new medicines.

What we found is that this type of collaborations may be relevant for countries in the Global South not only for updating their technological capabilities, but also to increase their visibility and impact in the scientific community of their field. Yet, it is relevant to acknowledge that co-authorships with the industry are far from being common in science, thus representing a case of corner outcomes with an edge at zero and a continuous distribution for strictly positive values (our sample as mean value of 1.73%). Our results show that, in line with Tijssen (2012), the intensity of science-industry co-authorship is lower in African and Latin American countries, than in countries in the North. Therefore we should be cautious when interpreting this result, because few publications in collaboration with industry can change substantially this indicator (high sensibility). Besides this, if we add to our model a variable that interacts industry collaboration intensity with international collaboration intensity, the covariate industry collaboration changes signal and the significance disappears. At the same time, the international collaboration parameter remains positive and significant. Therefore, it is not clear if the positive and significant effect of industry collaboration intensity in citation impact in our general model is due to the industry "effect" or is just because most industry collaborations are also international collaborations.

For countries that have English as an official language, our results show that the relation is positive and significant in almost every model specification. Since the majority of scientific journals are written in English and articles published in a non-English language have less potential readers, this positive relation was an expected result. In the Global South, an Anglophone colonial history and concomitant opportunities for partnerships with English speaking countries, for example by housing international research institutes, may have a significant effect on their citation impact.

These results also indicate that, contrary to what has been revealed by King (2004), the relation between GDPpc and citation impact is not strictly positive. It would be expectable that wealthier countries would have more resources to apply for science and therefore perform better in terms of citation impact. However, for countries in the Global South the coefficients are negative, and in the North they are positive and negative (non-significant) depending on the model specification. We also tried to understand if the U relationship showed in Fig. 2 and 3 holds in our model with all countries. Yet, when we include the variable GDPpc squared in the regression (see table 4 in appendix), the coefficient is positive but non-significant. These results indicate that there are other elements beyond wealth intensity that matter for research quality in the South, namely previous performance, a higher level of international collaboration and more publications in the specific subject area.

Finally, for country size, in terms of total scientific output there is no clear pattern since our results differ depending on the model specification. However, countries with higher population seem to have on average less citation impact than smaller countries. A possible interpretation for this is that smaller countries are more involved in international collaborations to produce their scientific articles. This may be so, as when we interact country size with level of international collaboration, the negative effect of population size is no longer significant and the interaction variable seems to capture that effect. For example, Frame & Carpenter (1979) also argued that the scientific size of a nation determines the need for international collaboration. Small countries have fewer opportunities to find collaborators inside their own country when compared to larger countries, thus having a greater need for research partners from other countries (Narin et al., 1991). Our results do not show significant difference between the South and North.

To complement this analysis, in appendix we do two different robustness checks. In table 5, to explore the performance of countries with different levels of international collaboration we create two sub-groups (i.e. low international collaboration intensity and high international collaboration intensity) in both Global South and Global North. In table 6, we do the same analysis as in table 3 but instead of separating the World in South and North, we use four broad World regions to see if there are significant differences between them. In general these results are consistent with the previous models. In table 5, we show that previous citation impact contributes more in the lower international collaboration group; while number of publications, level of international collaboration and level of collaboration with industry have a higher effect in the higher international collaboration group. In table 6, the main findings are that previous citation impact is more relevant in Africa and Latin America & Caribbean is the region where international collaboration has a higher effect on citation impact.

5. Discussion & Conclusion

In this article, we apply bibliometric and econometric analysis to identify which countries are producing research with higher scientific influence, and to understand which factors lead to those higher results. We focus specifically on the Global South because some of these countries are starting to converge in terms of output with the Global North and we found some evidence suggesting that at different levels of wealth intensity countries have different determinants of citation impact. We found that previous citation impact, level of international collaboration and publication output in a specific scientific field are important determinants of citation impact among all nations. The variable number of publications is the only factor that appears to be substantially more important in the South than in the North. This covariate represents not only the scientific output but also the intensity of involvement in scientific activities of a country in a specific area (as gross expenditure on R&D and number of researchers are usually highly correlated with number of publications). The agglomeration effects that may arise in some disciplines in scientific communities that are generally much smaller than their counterparts in the North seem therefore to be relevant. This implies that the importance of applying resources and generating a higher critical mass in a specific field, in order to produce research with more influence in the World, is greater in the South.

As for our lag dependent variable, we confirmed that it has an important effect on citation impact in both groups of countries. Societies vary in their capacity to produce major scientific discoveries over time because they are influenced in various ways by several historical processes and institutional environments. As a result of this sticky and tacit dimension of scientific knowledge, researchers in different types of organisations in the same country engage in a great deal of common learning and socialisation that is transmitted across time and organisations. This know-how is lost only if these individuals go away and don't interact with their local network. Consequently one potential implication for this result is that, because countries in the Global South have on average relatively few of these "excellent" researchers, "brain-drain" may have a worst effect on their scientific performance. If their few best "minds" go abroad to work or continue their studies, and don't comeback or interact with their national colleagues, then the tacit knowledge that they have and potential spillovers that they can generate will not be used for those countries gain.

Concerning level of international collaboration, as has been widely showed in the past, there is a positive and significant relationship with citation impact. With the advances in information and communication technology, and institutional changes, scientists can more easily obtain relevant knowledge by collaborating with other peers with diverse areas of expertise and backgrounds. Accessing external complementary knowledge and skills through networking, namely with scientists working in more developed environments, seems to be extremely significant for performing research with high impact. Yet, interestingly our results suggest that, contrary to what would be expected, this covariate does not seem more relevant in the South than in the North. This therefore indicates that the interest in pursuing international collaborations seems equally relevant for both groups of countries.

Our analysis also suggests that industry collaboration seems to be positively associated with citation impact, especially in the Global South. However, it is not clear if the positive and significant effect of industry collaboration intensity in citation impact in our general model is due to the "industry effect", or is just because the relatively few industry collaborations performed by the South are also international collaborations.

In our regressions we have also used country controls. We found that smaller countries (population wise) and countries with English as an official language perform on average better than others in some model specifications. A possible interpretation for this is that smaller countries rely further on international collaborations to produce their scientific articles. When we interact country size with level of international collaboration, the negative effect of population size is no longer significant and the interaction variable seems to capture that effect. For countries that have English as an official language, since the majority of scientific journals are written in English and articles published in a non-English language have less potential readers, this positive relation was an expected result. Besides this, countries that have English as an official language usually have a colonial legacy with Anglo-Saxon countries (US, UK, Canada, Australia) and consequently more collaborations with them (Pouris, 2010; Mêgnigbêto, 2013). These are the leading countries in many scientific fields therefore this positive relation is reinforced. Finally, there is no clear relation between wealth intensity (as measured by GDPpc) and citation impact. It would be expectable that wealthier countries would have more resources to apply for science and therefore perform better in terms of citation impact. However, we found that other elements beyond wealth intensity are much more relevant for the research quality of nations.

The paper also tries to explore potential biases in the use of bibliometric indicators in the Global South. It is widely accepted that these types of indicators capture poorly certain types of research and they encourage certain scientific activities and behaviours. For example, shift towards English publications (Hicks et al., 2015), diversion of research away from local or national issues (Hicks et al., 2015), scientific supply poorly aligned with societal needs (Sarewitz & Pielke, 2007), bias toward positive reporting (Fanelli, 2011), etc. As the Global South has a "lower" status in the scientific enterprise, these effects may be even worse for this group of countries. Another important issue when measuring citation impact is to be aware that it is a relative indicator. For example, if a country has the same citation impact (measured in intensity) as the US but it has 1000 times less publications than the US, evidently the actual absolute impact (scientific, societal and economical) of their research in the World is completely different. Therefore, even when using thresholds as we did, indicators measuring citation impact should be always interpreted within their context.

As this study was mainly carried out in a macro perspective, based on bibliometric indicators, there was some lack of understanding about the specificities of the national scientific systems. Complementing this quantitative analysis with a more qualitative approach (such as researching why specific institutions in the Global South have such high performance levels and understanding their interactions) would certainly improve the level of knowledge about science in lower income contexts. Also, to improve this model by using measures of network centrality instead of level of international collaboration, could give us a better understanding of the role of scientific network co-authorships for citation impact.

In what regards the normative implications, our findings allow drawing some potentially relevant indications. Lower and middle-income countries with globally small scientific communities would better concentrate their resources in generating higher critical masses in specific fields, in order to produce research with higher impact. Further, the interest in pursuing international collaborations seems more than justified, in case increasing scientific impact is an objective.

International scientific collaborations have been pursued more intensely by smaller countries, and this is comprehensible given the fact that larger countries may have larger numbers of researchers in every single major discipline, thus the need to collaborate abroad does not arise in the latter as in the former. However, even for the larger countries there may be good reasons to seek collaboration abroad at least in some fields, balancing this orientation without jeopardising the cohesion of their research systems.

These recommendations assume that increasing the impact of scientific publication in the South is an important objective, and that such impact is directly related to the quality of the produced research. However, it might be relevant to keep in mind the distinction between academic and practical impact. Though one may assume that in the long-term and globally both impacts may coincide, wise policy-makers in countries belonging to the Global South may recognize that for shorter time-spans and in specific geographic or institutional conditions that may not be so.

Finally the science policy-making process needs to keep in mind the strong path-dependency that dominates scientific activities globally. Despite the success stories of a few lower and middle-income countries that forged ahead in scientific matters in the most recent decades, most countries in the Global South are still held back by the chains of path-dependency. Overcoming such path-dependencies implies persistence, continuous investment and far-reaching institutional change, as those successful cases have confirmed.

6. Acknowledgements

The authors would like to thank Robin Cowan, Bart Verspagen, Johann Mouton and Pierre Mohnen for advice and insights. This paper also benefited from comments made by the participants at the workshop "Science and development: Growth, expansion and role of the university system in South Africa" (Stellenbosch, 2016).

7. Bibliography

Adams, J. (2013). Collaborations: The fourth age of research. *Nature*, *497*(7451), 557–560.

Akhondzadeh, S. (2013). Iranian science shows world's fastest growth: ranks 17th in science production in 2012. *Avicenna Journal of Medical Biotechnology*, *5*(3), 139.

Albuquerque, E. da M. e. (2004). Science and Technology Systems in Less Developed Countries. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 759–778). Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/1-4020-2755-9_35

Baldi, S. (1998). Normative versus Social Constructivist Processes in the Allocation of Citations: A Network-Analytic Model. *American Sociological Review*, *63*(6), 829–846.

Barnard, H., Cowan, R., & Müller, M. (2012). Global excellence at the expense of local diffusion, or a bridge between two worlds? Research in science and technology in the developing world. *Research Policy*, *41*(4), 756–769.

Bernardes, A. T., & Albuquerque, E. da M. e. (2003). Cross-over, thresholds, and interactions between science and technology: lessons for less-developed countries. *Research Policy*, *32*(5), 865–885.

Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80.

Bornmann, L., & Leydesdorff, L. (2013). Macro-Indicators of Citation Impacts of Six Prolific Countries: InCites Data and the Statistical Significance of Trends. *PLoS ONE*, *8*(2).

Bornmann, L., Schier, H., Marx, W., & Daniel, H.-D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics*, *6*(1), 11–18.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

Collazo-Reyes, F. (2014). Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country. *Scientometrics*, *98*(1), 197–209.

Confraria, H., & Godinho, M. M. (2014). The impact of African science: a bibliometric analysis. *Scientometrics*, *102*(2), 1241–1268.

DFID. (2014). *What is the Evidence on the Impact of Research on International Development?* London: UK Department for International Development.

Didegah, F., & Thelwall, M. (2013). Determinants of research citation impact in nanoscience and nanotechnology. *Journal of the American Society for Information Science and Technology*, *64*(5), 1055–1064.

Fagerberg, J., & Godinho, M. M. (2004). Innovation and catching-up. In *The Oxford Handbook of Innovation* (Fagerberg, J., Mowery, D., Nelson, R., pp. 514–543). Oxford: Oxford University Press. Retrieved from http://folk.uio.no/janf/downloadable_papers/03fagerberg-godinho_catching-up_ottawa.pdf

Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904.

Frame, J. D., & Carpenter, M. P. (1979). International Research Collaboration. *Social Studies of Science*, *9*(4), 481–497.

Garfield, E. (1979). *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. Isi Press.

Gilbert, G. N. (1977). Referencing as Persuasion. *Social Studies of Science*, *7*(1), 113–122.
Glänzel, W., Rinia, E. J., & Brocken, M. G. M. (1995). A Bibliometric Study on Highly-cited European Physics Papers in the 80s Based on "Physics Briefs" Source Publications and "SCI" Citation Data. *Research Evaluation*, *5*(2), 113–122.
Godin, B. (1996). Research and the practice of publication in industries. *Research Policy*, *25*(4), 587–606.
Godin, B., & Gingras, Y. (2000). What is scientific and technological culture and how is it measured? A multidimensional model. *Public Understanding of Science*, *9*(1), 43–58. http://doi.org/10.1088/0963-6625/9/1/303
Goldemberg, J. (1998). What Is the Role of Science in Developing Countries? *Science*, *279*(5354), 1140–1141.
Goldfinch, S., Dale, T., & Jr, K. D. (2003). Science from the periphery: Collaboration, networks and "Periphery Effects" in the citation of New Zealand Crown Research Institutes articles, 1995-2000. *Scientometrics*, *57*(3), 321–337.
Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431.
Hollingsworth, J. R. (2006). The dynamics of American Science : an institutional and organizational perspective on major discoveries. In *Transformationen des Kapitalismus : Festschrift für Wolfgang Streeck zum sechzigsten Geburtstag* (pp. 361–380). Frankfurt, New York: Campus Verlag.
InCites[TM]. (2014). *InCitesTM*. Thomson Reuters. Retrieved from https://incites.thomsonreuters.com
Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, *40*(3), 541–554.
King, D. A. (2004). The scientific impact of nations. *Nature*, *430*(6997), 311–316.
Knorr-Cetina, K. (1981). *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Pergamon Press.
Lall, S. (2000). Technological change and industrialization in the Asian newly industrializing economies : achievements and challenges. In L. Kim & R. R. Nelson (Eds.), *Tecnology, Learning and Innovation* (pp. 13–68). Cambridge: Cambridge University Press.
Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, Mass: Harvard University Press.
Latour, B., & Woolgar, S. (1979). *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press.
Lawrence, P. A. (2003). The politics of publication. *Nature*, *422*(6929), 259–261.
Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution*, *20*(1), 28–32.
Marx, W., & Bornmann, L. (2014). On the causes of subject-specific citation rates in Web of Science. *Scientometrics*, *102*(2), 1823–1827.
May, R. M. (1997). The Scientific Wealth of Nations. *Science*, *275*(5301), 793–796.
Mazzoleni, R., & Nelson, R. R. (2007). Public research institutions and economic catch-up. *Research Policy*, *36*(10), 1512–1528.

McCaffrey, D. F., Mihaly, K., Lockwood, J. R., & Sass, T. R. (2012). A review of Stata commands for fixed-effects estimation in normal linear models. *Stata Journal*, (12), 406–432.

Mêgnigbêto, E. (2013). International collaboration in scientific publishing: the case of West Africa (2001–2010). *Scientometrics*, *96*(3), 761–783.

Merton, R. K. (1968). The Matthew Effect in Science The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.

Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Moed, H. F. (2005). *Citation Analysis in Research Evaluation*. Springer Science & Business Media.

Moed, H. F., de Moya-Anegón, F., López-Illescas, C., & Visser, M. (2011). Is concentration of university research associated with better research performance? *Journal of Informetrics*, *5*(4), 649–658.

Moulton, B. R. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit. *The Review of Economics and Statistics*, *72*(2), 334–38.

Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*, *21*(3), 313–323.

Nelson, R. R. (2005). The Roles of Research in Universities and Public Labs in Economic Catch-up. In G. Santangelo, *Technological Change and Economic Catch-up*. Edward Elgar Publishing. Retrieved from http://www.elgaronline.com/view/9781845424824.00009.xml

Peters, H. P. F., & van Raan, A. F. J. (1994). On determinants of citation scores: A case study in chemical engineering. *Journal of the American Society for Information Science*, *45*(1), 39–49.

Pouris, A. (2010). A scientometric assessment of the Southern Africa Development Community: science in the tip of Africa. *Scientometrics*, *85*(1), 145–154.

Pudovkin, A. I., & Garfield, E. (2009). Percentile rank and author superiority indexes for evaluating individual journal articles and the author's overall citation performance. *CollNet Journal of Scientometrics and Information Management*, *3*(2), 3–10.

Puuska, H.-M., Muhonen, R., & Leino, Y. (2013). International and domestic co-publishing and their citation impact in different disciplines. *Scientometrics*, *98*(2), 823–839.

Raan, A. F. J. van. (2004). Measuring Science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 19–50). Springer Netherlands. Retrieved from http://link.springer.com/chapter/10.1007/1-4020-2755-9_2

Radosevic, S., & Yoruk, E. (2014). Are there global shifts in the world science base? Analysing the catching up and falling behind of world regions. *Scientometrics*, *101*(3), 1897–1924.

Riviera, E. (2015). Testing the strength of the normative approach in citation theory through relational bibliometrics: The case of italian sociology. *Journal of the Association for Information Science and Technology*, *66*(6), 1178–1188.

Salter, A. J., & Martin, B. (2001). The economic benefits of publicly funded basic research: a critical review. *Research Policy*, *30*(3), 509–532.

StataCorp. (2013). Stata Statistical Software (Version 13). College Station, TX: StataCorp LP.

Testa, J. (2011). *The Globalization of Web of Science: 2005-2010*. Philadelphia: Thomson Reuters. Retrieved from http://wokinfo.com/media/pdf/globalwos-essay.pdf

Tijssen, R. J. W. (2012). Co-authored research publications and strategic analysis of public–private collaboration. *Research Evaluation*, *21*(3), 204–215.

Tijssen, R. J. W., Leeuwen, T. N. van, & Wijk, E. van. (2009). Benchmarking university-industry research cooperation worldwide: performance measurements and indicators based on co-authorship data for the world's largest universities. *Research Evaluation*, *18*(1), 13–24.

UNESCO. (2010). *UNESCO science report 2010*. Paris: United Nations Educational, Scientific and Cultural Organization. Retrieved from http://www.unesco.org/science/psd/publications/usr_2010.pdf

UNESCO (Ed.). (2015). *UNESCO science report: towards 2030*. Paris: UNESCO Publishing.

Van Raan, A. F. (1998). The influence of international collaboration on the impact of research results. *Scientometrics*, *42*(3), 423–428.

Waltman, L., & Eck, N. J. van. (2012). Source normalized indicators of citation impact: an overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, *5*(1), 37–47.

White, H. D. (2004). Reward, persuasion, and the Sokal Hoax: A study in citation identities. *Scientometrics*, *60*(1), 93–120.

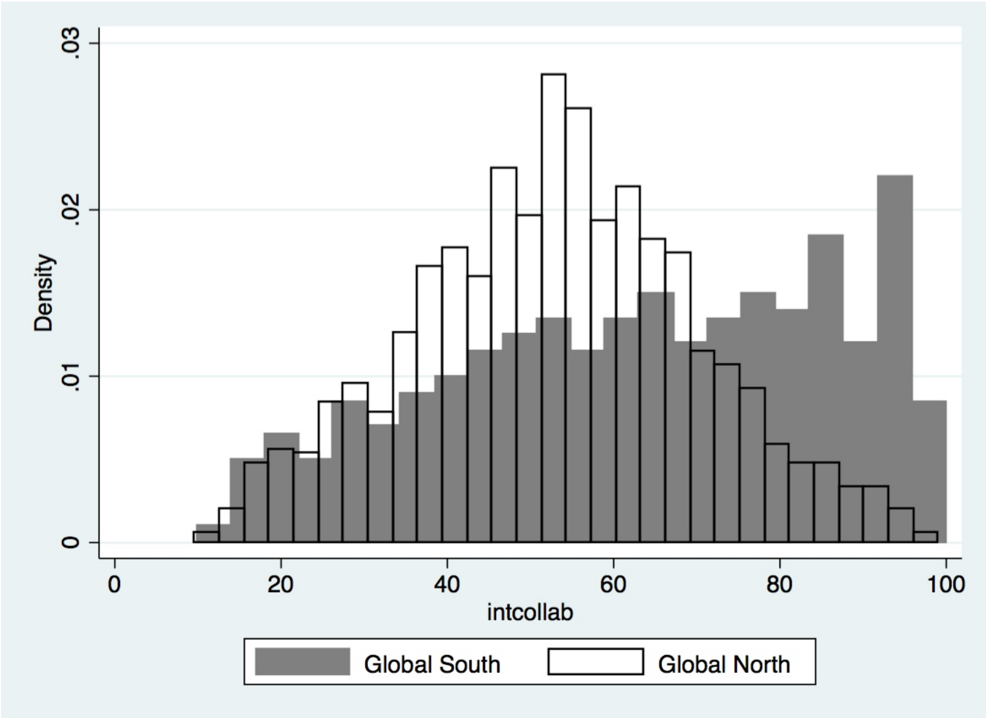
8. Appendix

Table 1. Significant determinants of citation impact based on previous studies (not exhaustive)

Level and Factors	What associates with higher citation	Prior literature
Article		
Number of authors	Four or more authors	Leimu & Koricheva, 2005
Length of the abstract	Longer abstract	Leimu & Koricheva, 2005
Journal impact factor (JIF)	Articles in journals with higher JIF	Peters & van Raan, 1994; Didegah & Thelwall, 2013
Number of references	More references	Peters & van Raan, 1994
Impact of references	Higher no. of citations	Bornmann et al., 2012; Didegah & Thelwall, 2013
Length of the paper	Longer paper	Peters & van Raan, 1994
Type of document	Reviews	Peters & van Raan, 1994
Language	English journal and paper	Peters & van Raan, 1994
Author		
Country of origin	Native English-speaking authors	Leimu & Koricheva, 2005
Previous performance	More citations in the past	Merton, 1968
Institution		
Size	Universities with a large publication output	Moed et al., 2011
Number of institutions	More institutions	Narin et al., 1991
Specialization intensity	Weak negative effect	Moed et al., 2011
Country		
Economic development	Higher GDP per capita	King, 2004
Number of countries of affiliation	More countries	Narin et al., 1991; Glänzel et al., 1995; Katz & Hicks, 1997; Van Raan, 1998; Puuska et al., 2013
Country of affiliation	English speaking country	Leimu & Koricheva, 2005

Source: Own elaboration.

Fig. 4. Distribution of international collaboration levels. South vs North (2008-2012)



Source: Own calculations based on *InCites*™
Note: Vertical axis shows the density of observations in a specific level of international collaboration in 2008-2012; Horizontal axis shows level of international collaboration.

Table 4. Determinants of citation impact in all countries

Variables	**PPtop10%_0812**	**CXC_0812**
PPtop10%_0307	0.590***	
	(0.040)	
CXC_0307		0.606***
		(0.060)
Pubs area (log)	1.507***	0.066*
	(0.324)	(0.033)
Int. Collab	0.085***	0.007***
	(0.008)	(0.001)
Ind. Collab	0.214**	0.017***
	(0.084)	(0.006)
GDPpc (log)	-2.867	-0.350*
	(2.140)	(0.205)
GDPpc^2 (log)	0.378	0.038
	(0.291)	(0.027)
English Official	0.910***	0.053**
	(0.254)	(0.021)
Total Pubs (log)	0.803**	0.134***
	(0.335)	(0.037)
Popul. (log)	-1.451***	-0.137***
	(0.330)	(0.029)
Constant	6.764	0.992**
	(4.486)	(0.436)
Observations	1,686	1,686
R-squared	0.726	0.638

Note 1: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
Note 2: Linear regression, absorbing indicators. Std. error adjusted for 126 clusters (countries).

Table 5. Determinants of citation impact in the Global South and the Global North by international collaboration groups

Variables	PPtop10%_0812			
	South (GDPpc)		**North (GDPpc)**	
	Int. Collab >= average (56%)	Int. Collab< average (56%)	Int. Collab>= average (56%)	Int. Collab< average (56%)
PPtop10%_0307	0.482***	0.526***	0.663***	0.688***
	(0.074)	(0.066)	(0.094)	(0.049)
Pubs Area (log)	2.461***	1.188*	0.322	0.495
	(0.924)	(0.622)	(0.432)	(0.391)
Int. Collab	0.125***	0.153***	0.056***	0.047***
	(0.023)	(0.024)	(0.017)	(0.012)
Ind. Collab	0.325**	0.249**	0.098	0.052
	(0.129)	(0.103)	(0.204)	(0.081)
Total Pubs (log)	0.031	2.339***	1.042*	0.641
	(0.667)	(0.730)	(0.568)	(0.528)
English Official	1.510**	0.758*	-0.327	0.674**
	-0.58	-0.387	-0.46	-0.307
GDPpc (log)	0.208	-0.307	-0.278	2.064***
	(0.744)	(0.696)	(0.854)	(0.658)
Popul. (log)	-1.504**	-2.292***	0.096	-0.956**
	(0.632)	(0.567)	(0.428)	(0.426)
Constant	-1.069		-5.708	
	(3.627)		(3.564)	
Observations	815		871	
R-squared	0.682		0.805	

Note 1: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
Note 2: Linear regression, absorbing indicators. Std. error adjusted for 126 clusters (countries).

Table 6. Determinants of citation impact in four World regions (Africa; Asia; LA&C - Latin America & Caribbean; E&NA&P - Europe & North America & Pacific)

	Dependent variables							
	PPtop10%_0812				CXC_0812			
Variables	Africa	LA&C	Asia	E&NA&P	Africa	LA&C	Asia	E&NA&P
PPtop10%_0307	0.690***	0.477***	0.532***	0.598***				
	(0.062)	(0.104)	(0.077)	(0.052)				
CXC_0307					0.811***	0.354**	0.558***	0.602***
					(0.171)	(0.144)	(0.068)	(0.084)
Pubs Area (log)	1.139**	2.349***	1.888***	1.494***	0.098*	0.197***	0.074*	0.052
	(0.523)	(0.490)	(0.524)	(0.425)	(0.056)	(0.064)	(0.044)	(0.048)
Int. Collab	0.069***	0.122***	0.069***	0.106***	0.004**	0.013***	0.006***	0.009***
	(0.012)	(0.025)	(0.014)	(0.012)	(0.002)	(0.003)	(0.001)	(0.001)
Ind. Collab	0.353***	-0.097	0.109	0.252**	0.057	0.005	0.009	0.016**
	(0.111)	(0.081)	(0.078)	(0.100)	(0.036)	(0.006)	(0.006)	(0.007)
Total Pubs (log)	-0.610	-0.040	0.341	0.110	-0.062	0.034	0.122***	0.102
	(0.551)	(0.669)	(0.461)	(0.697)	(0.106)	(0.055)	(0.043)	(0.077)
English Official	0.478	0.340	1.193**	1.120***	0.033	0.085	0.049	0.080***
	(0.401)	(0.871)	(0.509)	(0.374)	(0.066)	(0.082)	(0.035)	(0.026)
GDPpc (log)	-0.351	-0.892	0.571	0.374	-0.092**	-0.263**	-0.023	-0.037
	(0.650)	(0.785)	(0.406)	(0.577)	(0.043)	(0.132)	(0.035)	(0.054)
Popul. (log)	0.028	-0.566	-0.996**	-0.925*	-0.007	-0.013	-0.104***	-0.107**
	(0.401)	(0.423)	(0.411)	(0.511)	(0.068)	(0.071)	(0.034)	(0.051)
Constant	-1.814				0.233			
	(3.051)				(0.274)			
Observations	1,686				1,686			
R-squared	0.740				0.661			

Note 1: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
Note 2: Linear regression, absorbing indicators. Std. Err. adjusted for 126 clusters (countries).
Note 3: Africa = 221 observations (28 countries); LA&C = 195 observations (16 countries); Asia = 329 observations (34 countries); E&NA&P = 832 observations (48 countries).

The UNU-MERIT Working Paper Series

2016-01 *Mexican manufacturing and its integration into global value chains* by Juan Carlos Castillo and Adam Szirmai

2016-02 *New variables for vocational secondary schooling: Patterns around the world from 1950-2010* by Alison Cathles

2016-03 *Institutional factors and people's preferences in social protection* by Franziska Gassmann, Pierre Mohnen & Vincenzo Vinci

2016-04 *A semi-endogenous growth model for developing countries with public factors, imported capital goods, and limited export demand* by Jan Simon Hallonsten and Thomas Ziesemer

2016-05 *Critical raw material strategies in different world regions* by Eva Barteková and René Kemp

2016-06 *On the value of foreign PhDs in the developing world: Training versus selection effects* by Helena Barnard, Robin Cowan and Moritz Müller

2016-07 Rejected Afghan asylum seekers in the Netherlands: Migration experiences, current situations and future aspirations

2016-08 *Determinants of innovation in Croatian SMEs: Comparison of service and manufacturing firms* by Ljiljana Bozic and Pierre Mohnen

2016-09 *Aid, institutions and economic growth: Heterogeneous parameters and heterogeneous donors* by Hassen Abda Wakoy

2016-10 *On the optimum timing of the global carbon-transition under conditions of extreme weather-related damages: further green paradoxical results* by Adriaan van Zon

2016-11 *Inclusive labour market: A role for a job guarantee* scheme by Saskia Klosse and Joan Muysken

2016-12 *Management standard certification and firm productivity: micro-evidence from Africa* by Micheline Goedhuys and Pierre Mohnen

2016-13 *The role of technological trajectories in catching-up-based development: An application to energy efficiency technologies* by Sheng Zhong and Bart Verspagen

2016-14 *The dynamics of vehicle energy efficiency: Evidence from the Massachusetts Vehicle Census* by Sheng Zhong

2016-15 *Structural decompositions of energy consumption, energy intensity, emissions and emission intensity - A sectoral perspective: empirical evidence from WIOD over 1995 to 2009* by Sheng Zhong

2016-16 *Structural transformation in Brazil, Russia, India, China and South Africa (BRICS)* by Wim Naudé, Adam Szirmai and Nobuya Haraguchi

2016-17 *Technological Innovation Systems and the wider context: A framework for developing countries* by Hans-Erik Edsand

2016-18 *Migration, occupation and education: Evidence from Ghana* by Clotilde Mahé and Wim Naudé

2016-19 *The impact of ex-ante subsidies to researchers on researcher's productivity: Evidence from a developing country* by Diego Aboal and Ezequiel Tacsir

2016-20 *Multinational enterprises and economic development in host countries: What we know and what we don't know* by Rajneesh Narula and André Pineli

2016-21 *International standards certification, institutional voids and exports from developing country firms* by Micheline Goedhuys and Leo Sleuwaegen

2016-22 *Public policy and mental health: What we can learn from the HIV movement* by David Scheerer, Zina Nimeh and Stefan Weinmann

2016-23 *A new indicator for innovation clusters* by George Christopoulos and Rene Wintjes

2016-24 *Including excluded groups: The slow racial transformation of the South African university system* by Helena Barnard, Robin Cowan, Alan Kirman and Moritz Müller

2016-25 *Fading hope and the rise in inequality in the United States* by Jo Ritzen and Klaus F. Zimmermann

2016-26 *Globalisation, technology and the labour market: A microeconometric analysis for Turkey* by Elena Meschi, Erol Taymaz and Marco Vivarelli

2016-27 *The affordability of the Sustainable Development Goals: A myth or reality?* By Patima Chongcharoentanawat, Kaleab Kebede Haile, Bart Kleine Deters, Tamara Antoinette Kool and Victor Osei Kwadwo

2016-28 *Mimetic behaviour and institutional persistence: a two-armed bandit experiment* by Stefania Innocenti and Robin Cowan

2016-29 *Determinants of citation impact: A comparative analysis of the Global South versus the Global North* by Hugo Confraria, Manuel Mira Godinho and Lili Wang