SPRU
Science Policy Research Unit

Working Paper Series

SWPS 2018-12 (July)

The Relation between Research Priorities and Societal Demands: The Case of Rice

Tommaso Ciarli and Ismael Ràfols

UNIVERSITY OF SUSSEX

SPRU Working Paper Series (ISSN 2057-6668)

The SPRU Working Paper Series aims to accelerate the public availability of the research undertaken by SPRU-associated people, and other research that is of considerable interest within SPRU, providing access to early copies of SPRU research.

Editors		**Contact**
Tommaso Ciarli		*T.Ciarli@sussex.ac.uk*
Daniele Rotolo		*D.Rotolo@sussex.ac.uk*

Associate Editors	**Area**	
Karoline Rogge	Energy	*K.Rogge@sussex.ac.uk*
Paul Nightingale, Ben Martin, & Ohid Yaqub	Science, & Technology Policy	*P.Nightingale@sussex.ac.uk* *B.Martin@sussex.ac.uk* *O.Yaqub@sussex.ac.uk*
Tommaso Ciarli	Development	*T.Ciarli@sussex.ac.uk*
Joe Tidd & Carlos Sato	Technology Innovation Management	*J.Tidd@sussex.ac.uk* *C.E.Y.Sato@sussex.ac.uk*
Maria Savona	Economics of Technological Change	*M.Savona@sussex.ac.uk*
Andrew Stirling	Transitions	*A.C.Stirling@sussex.ac.uk*
Caitriona McLeish	Civil Military Interface	*C.A.McLeish@sussex.ac.uk*

Editorial Assistance		
Martha Bloom		*M.Bloom@sussex.ac.uk*

Guidelines for authors

Papers should be submitted to swps@sussex.ac.uk as a PDF or Word file. The first page should include: title, abstract, keywords, and authors' names and affiliations. The paper will be considered for publication by an Associate Editor, who may ask two referees to provide a light review. We aim to send referee reports within three weeks from submission. Authors may be requested to submit a revised version of the paper with a reply to the referees' comments to swps@sussex.ac.uk. The Editors make the final decision on the inclusion of the paper in the series. When submitting, the authors should indicate if the paper has already undergone peer-review (in other series, journals, or books), in which case the Editors may decide to skip the review process. Once the paper is included in the SWPS, the authors maintain the copyright.

Websites

UoS: www.sussex.ac.uk/spru/research/swps

SSRN: http://www.ssrn.com/link/SPRU-RES.html

IDEAS: ideas.repec.org/s/sru/ssewps.html

Research Gate: www.researchgate.net/journal/2057-6668_SPRU_Working_Paper_Series

The Relation between Research Priorities and Societal Demands: The Case of Rice*

Tommaso Ciarli[†1] and Ismael Ràfols[‡2,3,1]

[1]SPRU, University of Sussex, UK
[2]*Ingenio* (CSIC-UPV), Universitat Politècnica València, Spain
[3]CWTS, University of Leiden, The Netherlands

July 17, 2018
Forthcoming in ***Research Policy***

Abstract

To what extent is scientific research related to societal needs? To answer this crucial question systematically we need to contrast indicators of research priorities with indicators of societal needs. We focus on rice research and technology between 1983 and 2012. We combine quantitative methods that allow investigation of the relation between 'revealed' research priorities and 'revealed' societal demands, measured respectively by research output (publications) and national accounts of rice use and farmers' and consumers' rice related needs. We employ new bibliometric data, methods and indicators to identify countries' main rice research topics (priorities) from publications. For a panel of countries, we estimate the relation between revealed research priorities and revealed demands. We find that, across countries and time, societal demands explain a country's research trajectory to a limited extent. Some research priorities are nicely aligned to societal demands, confirming that science is partly related to societal needs. However, we find a relevant number of misalignments between the focus of rice research and revealed demands, crucially related to human consumption and nutrition. We discuss some implications for research policy.

Keywords: Research priority; agenda setting; research trajectories; societal needs; scientometrics; rice

*This paper has benefited from discussions with several colleagues at different stages of the analysis and writing: Frederique Bone, Sally Brooks, Ramon Carreres Ortells, Jose Falck-Zepeda, Dominic Glover, Pierre Benoit Joly, Erik Millstone, David Spielman, Andy Stirling, Jim Sumberg, Ludovic Temple, Matthew Wallace, Ohid Yaqub; and from discussions during the SPRU seminars (2017 & 2014), S&T indicators (2016), SPRU 50 (2016), 4S at Barcelona (2016), Eurkind at Ingenio (2016), contested agronomy at IDS (2016), Atlanta Science Policy (2015), EMAEE at Maastricht (2015), 4S at Buenos Aires (2014), GRC conference (2014), Global TechMining Conference at Atlanta (2013). Comments from two anonymous reviewers and Ed Steinmueller, co-editor of this Issue, and two anonymous reviewers and Daniele Rotolo, editor of the SPRU Working Paper Series helped to substantially improve the paper. We are grateful to Cristina Ashby at CABI for providing the organised publication data, to Alan Porter for providing access to Vantage Point for initial data cleaning, and to Ludo Waltman for discussing tweaks and tricks in VosViewer. The paper received funding from the UK ESRC project RES-360-25-0076 and the US National Science Foundation (Award # 1064146). The findings and observations in this paper are the authors' and do not necessarily reflect the views of the NSF.

†SPRU, Jubilee Building, University of Sussex, BN1 9SL Falmer, Brighton, UK, T.Ciarli@sussex.ac.uk
‡i.rafols@ingenio.upv.es

1 Introduction

In recent years there has been an incipient shift from evaluation systems focused on academic excellence to systems that take account also of 'societal impact'. This shift is particularly relevant in mission-oriented areas, such as health and agriculture (Wright, 2012), where research is related explicitly to social goals such as reducing the disease burden or improving yields (Kahlon et al., 1977; Joly et al., 2015). However, some authors suggest that evaluation should consider not only the magnitude of the social impact of research but also the type of impact, for example, whether and to what extent it addresses and satisfies societal needs (Pinstrup-Andersen and Franklin, 1977; Bantilan and Keatinge, 2007; Sarewitz and Pielke, 2007; Kinge et al., 2014). "For the effective allocation of their scarce human and financial resources, institutions such as those involved in public agricultural research must take into consideration the needs of farmers as well as overall national, social, and economic goals" (Pinstrup-Andersen and Franklin, 1977, p. 416).

In this paper we study whether agricultural research priorities are and agriculture related societal demands (or needs) for research are related across countries during the period 1983-2012. We study this relation from an aggregate perspective, proxying relative focus on a research priority by research output (publications), and societal demands by country use of its agriculture production (e.g., export, food, seeds, etc.) and basic societal needs (e.g., nutrition and productivity). Following Sarewitz and Pielke (2007), we refer to a close relation between priorities and needs as 'alignment', and a lack of relation as 'misalignment', – without however implying that there is a unique best formula for prioritisation.

We are not aiming to provide an evaluation of the impact of rice research on farmers' output, consumers' health or any other social indicator (e.g., nutrition). As many authors argue (e.g. Dalrymple, 1977; Pardey et al., 2016), observing the impact of agriculture research on specific outcomes takes time, and the impact involves a large number of processes, which, often, are difficult to identify. The attribution of societal impact to specific research efforts is problematic (Matt et al., 2017; Spaapen and van Drooge, 2011). Rather, we investigate whether research prioritised in a country during a given time period focuses on the problems perceived as important in that same period. Our paper relates to work done by Evans et al. (2014) and Rafols and Yegros (2018) on health research and disease burden. In this paper, we focus on production and use of rice.

For instance, low rice yields may be due to the country's use of inputs, type of rice cultivated, cultivation ecosystem, water availability, cultivation practices, etc. Low yields may be a problem for the farmer (low value added) and for the consumer (relative high price of rice). To simplify, let us assume that low yields have a greater impact on the farmer needs because consumers have access to other food crops and may not rely mainly on rice. If low yield is perceived as a more pressing problem than, for example, seed availability, access to calories or pests, we may expect an increase in the demand for research to improve rice yields. The detailed focus of such research might be dependent on which yields related aspects are considered the most relevant and urgent. Conventional research impact evaluation would assess whether some previous research projects or programmes on rice yields have achieved or have the potential to achieve societal impact in terms of improved yields. Our interest is in understanding whether a relative increase in domestic demand for research related to rice yields has resulted in an increase in the relative share of research related to yield.

Rice is an interesting case study because of its socio-economic importance. It is a basic good that has been produced for centuries, which incorporates high technology components with centuries of knowledge and experience and which satisfies multiple needs (e.g., it feeds 3 bn people) (Woolston, 2014). Rice is also at the core of the green revolution, a huge west-led, world-wide investment in innovation in agriculture research and technology. So diverse has been the impact of the green revolution across countries, regions, farmers and data sources (Hsieh and Ruttan, 2015; Orr, 2012; Evenson and Gollin, 2003), that it is plausible to assume that research was unevenly aligned to needs and demands, which differ across regions and actors. Rice research and technologies continue to be controversial research areas (Eisenstein, 2014), which are likely to be shaped by different factors, in different locations, cultures and political economies. Our results cannot be directly extrapolated to other research areas and societal issues, but our study provides a methodology and a benchmark for the application of similar quantitative methods to other research controversies (including other crops).

We first applied co-word similarity to the published research on rice, and clustered similar terms into topics that identify different areas of research on rice. The topics identified represent the areas on which different actors in the agriculture research system have published. Subject to resource constraints, research organisations make decisions about which areas to prioritise and how to distribute their research efforts (investment) across a range of topics – a research portfolio (Wallace and Rafols, 2015). Lack of reliable information on the funding of rice research by field and by country, makes publications the best available proxy to reveal organisations' priorities.[1] The clustering identifies six main rice research topics between 1983 and 2012: (i) plant protection, divided into pests and weeds; (ii) plant nutrition and yields; (iii) rice varieties and classical genetics; (iv) transgenic, molecular biology and genomics; (v) consumption, human nutrition and food technologies; and (vi) production practices and socio-economic issues.

We measure revealed national research priorities using the share of publications by topic, country and year for the main rice producing and/or publishing countries. We also measure countries' revealed demands using data on rice demand across seven main aggregates – exports, direct food consumption, seeds, food processing industry, imports, waste and animal feed – and on the main needs of farmers (e.g., use of fertilisers and pesticides) and consumers (e.g., rice caloric intake).

We find that the distribution of published research across research topics varies signif-

[1] On average, there is a linear relation between funding and publications, as shown in the appendix for health and agriculture research in the US, for which we have reliable data on funding and publications.

icantly across countries and time. Topics such as plant protection, socio-economic issues and, recently, traditional genetics, have become less relevant, while genomics and human nutrition have attracted conspicuous and growing amounts of research. Within these average tendencies, countries have followed different trajectories. Exploiting the variation in the shares of publications by country and year across topics, we estimate the relation between countries' demands and priorities, for a panel of 16 countries, over 28 years.

We find that across countries and time, revealed demand for rice technologies, to a limited extent, explain a country's publications distribution across topics, i.e. the 'revealed' research priorities. In some cases, the alignment between demands and priorities is as expected. For instance, countries with low yields invest more in plant nutrition, traditional genetics and production practices. However, we also find a large number of misalignments between the focus of rice research and revealed demands. For instance, food/nutrition needs are not met by increased specialisation in research on human nutrition. Rather, countries where rice represents an important source of caloric intake, specialise significantly less in food technologies, traditional genetics and genomics (although this effect disappears if we lag the revealed demand). Also worrying is the finding that, genomics, currently the research field with most publications, is the least related to demand for rice technology.

The paper makes several contributions. First, we contribute to the literature on research priority setting (e.g. Ely et al., 2013; Wallace and Rafols, 2016; Gläser and Laudel, 2016), especially work on agriculture (e.g. Arvanitis and Chatelin, 1988; Norton et al., 1992; Kelley et al., 1995; Sumberg, 2002; Dalrymple, 2006; Bantilan and Keatinge, 2007; Raitzer and Norton, 2009; Vanloqueren and Baret, 2009; Touzard and Temple, 2012; Sumberg et al., 2013).

From a methodological point of view, we develop quantitative methods to identify and visualise topics in large publication corpora and to compare the distribution of publications across research topics, to the salience of socio-economic issues such as societal demand. There is little quantitative evidence on how the direction of science and technology in one sector, for a specific technology, is influenced by policy or socio-economic factors (Gläser and Laudel, 2016). This is, in part, because scholarly investigation of the relationship between research priorities, technological trajectories and societal needs faces major methodological difficulties. We identify three main challenges. First, mapping past and current outputs of research investments in terms of topics, and how these topics change through time. Second, measuring some societal needs across time and measuring their relative importance. Third, and possibly the most difficult, investigating the relation between the distribution of research priorities and the distribution of societal needs. We try to address these issues by investigating the factors influencing research priorities. We hope this work encourages future systematic quantitative examination of how different factors influence research priorities in science and technology, and their trajectories.

From a science policy perspective, our study is a first attempt to identify priorities in rice research in the long run and estimate their alignment to societal demands related to rice production and consumption. To our knowledge, ours is the first study to estimate the relation between research priorities and societal demands across countries for a given technology.[2] One study that explores methods and topics to study this relation at the world level is (Cassi et al., 2017). We believe that the methodology we propose could help national Science and Technology organisations to reflect on their past and current research portfolios on crucial technologies. To our knowledge, there is only one paper that provides a comprehensive investigation of rice research using publication data; Morooka et al. (2014) find that rice research in Japan cuts across a large number of disciplines, shifting between the 1970s and 2000s, from agriculture- topics to biology-related topics.

Second, we reveal a number of misalignments between research trajectories and societal demands, which might help reflection on the relation between demand and supply of science Sarewitz and Pielke (2007); McNie (2007); Bozeman and Sarewitz (2011); Wallace and Rafols (2015). Although the literature has discussed several examples, such as the case of research on fodder in Sub Saharan Africa (SSA), which has seen minimal rates of adoption among farmers (Sumberg, 2002), we believe that our systematic examination provides a broader perspective on these misalignments and may help to inform national agriculture research policies, especially for the case of rice.

Third, we contribute to exploring the shaping of technological trajectories by broad societal needs. We seek to reconcile insights from two main approaches to studying the direction of technology: that focused mainly on the properties of technologies and the constraints of technological paradigms and regimes, common in the economics of innovation (Verspagen, 2007; Jenkins and Floyd, 2001, e.g.); and that focused mainly on how social agencies and structures shape technology choices, common in social studies of technology (e.g., Pinch and Bijker, 1984; MacKenzie, 1998). We provide a preliminary attempt to combine these different explanations.

Section 2 discusses the background and motivation of the paper. Section 3 discusses our theoretical framework and Section 4 describes the data, and empirical methods used for their analysis. We present and discuss the results in Section 5, followed by Section 6, which concludes.

2 Background and Motivation

The direction of science and technology is influenced by a number of factors. In a seminal article, Dosi (1982) defines a technology trajectory within a given technological (and scientific) paradigm on the basis of technical and engineering factors.[3] Subsequent contributions from different social sciences, such as economics, sociology, management and history, focus on various (sets of) factors. The evolution of scientific and technological

[2] Evans et al. (2014) study prioritisation across diseases, i.e. across different problems.

[3] "the pattern of normal problem solving activity (i.e. of progress) on the ground of a technological paradigm. [...] Once a path has been selected and established, it shows a momentum of its own [...], which contributes to define the directions towards which the problem solving activity moves." (Dosi, 1982, 152-153)

trajectories is discussed in association to a number of techno-economic factors (Freeman, 1991), actors (Freeman, 1995), socio-economic factors (Dosi and Nelson, 2013; Smith et al., 2005) and political factors (Johnstone and Stirling, 2015). As suggested by (Pavitt, 1998, p. 793), "the rate and direction of the development of a country's science base is strongly influenced by its level of economic development, and the composition of its economic and social activities. In other words, it is socially shaped". However, the extent to which such diverse factors contribute to shaping the trajectory of science and technology is largely unknown. This is due partly to disciplinary barriers, partly to external validity – the relevance of each factor weights is likely to be sector, technology and region specific – and partly to methodological difficulties related to identifying empirically the roles of different factors.

Research priorities and societal needs are important for shaping the trajectory of scientific research and technology. However, it would seem that scientific advances are unevenly distributed across societal sectors and their diverse demands (de Janvry, 1978; Gibbons et al., 1994; Nelson, 2003, 2011). For example, there is low relative investment in research on diseases affecting poor populations (neglected diseases) compared to diseases affecting the global north, and in research on healthy lifestyles compared to the level of investment in pharmaceuticals research (Evans et al., 2014; Rotolo et al., 2013). Many research areas not included in public research agendas are pursued by civil society organisations (Frickel et al., 2009; Hess et al., 2017). Technological developments in agriculture often privilege specific forms of productivity at the expense of sustainability (Vanloqueren and Baret, 2008, 2009; Carlisle and Miles, 2016) and can tend to neglect local needs (Dalrymple, 2006).

The rhetoric on societal needs does not always match patterns in science and innovation (Barke and Jenkins-Smith, 1993; Plutzer et al., 1998). This reinforces the tendency for scientific and technological progress to follow directions driven only partly by societal needs (David and Sanderson, 1997; Miller and Neff, 2013; Mokyr, 2000).

The gap between research priorities and societal needs may be due to several socio-economic and political factors (Sarewitz, 2010). For example, key science policy priorities may be driven by problem framings within the scientific community (Bozeman and Sarewitz, 2005), as suggested by Bush (1945) and Polanyi (2000), rather than in response to wider societal demand. There may also be inequality in the distribution of resources and power: different actors (such as private companies, academia, mission-oriented laboratories) have different interests in developing science and innovation, and may invest in different areas within a broad research landscape (Chataway et al., 2004; Wallace and Rafols, 2016). For instance, international companies priorities in developing seed technologies may differ from those of local companiesShiva and Crompton (1998); Marin et al. (2014). More generally, international organisations may have a strong influence on domestic research, through research collaborations or because they push their own agendas Evenson and Gollin (2003); Wallace and Rafols (2016). Brooks (2011a) documents how international agriculture research can reduce the ability of low income countries to address their main needs.

Also, the majority of society plays a minor role in the direction of research and technologies even in societies with urgent primary needs such as nutrition and health. For example, Evans et al. (2014) discuss how health research publications are misaligned to the disease burden (measured by disability-adjusted life years). Klerkx and Leeuwis (2008a) uses case study evidence to suggest that in agriculture, farmers generally play a small role in shaping research and development funding.[4]

Finally, another explanation for why science and technology may evolve in directions that have a lower impact on social welfare or in directions not aligned to the demands of end users of science and technologies, may be path dependency. Lock-in to suboptimal trajectories can occur for a number of reasons such as economies of scale, network externalities, incumbent infrastructures, sunk costs or imitation. There are a number of examples in agriculture. David (1971), for the US Midwest, shows how economic conditions , such as factor prices and land size, slowed adoption of a new technology: the reaping machine. Cowan and Gunby (1996) shows that new, less polluting and less expensive Integrated Pest Management (IPM) technologies failed to replace chemical pesticides due to a combination of technological externalities, learning costs and uncertainty. Wolff and Recke (2000) provide further evidence of tomato farmers being locked in chemical pesticides in Ghana, which blocked the diffusion of the superior IPM. Vanloqueren and Baret (2008) show that a number of systemic factors, including lock-in to supplying companies, agricultural policies and lock-in of the extension services to a few standardised and well appraised technologies, delayed the adoption of disease resistant cultivars in Belgium. McGuire (2008) shows for the case of Ethiopia, that sunk costs in investments and learning shaped the breeding of modern varieties of sorghum. Hogg (2000) studied genetic diversity and the selection of new breeds and found they were shaped (locked-in) by policies, farming practices, available germplasms and established research routines.

Vanloqueren and Baret (2009) go beyond path dependency and examine how a number of systemic determinants (aspects related to the agricultural system of innovation) influenced the choice of one technological trajectory (genetic engineering) over a competing technology (agroecology). Among the determinants, they discuss agriculture science policies (and their interaction with the private sector and different interest groups), how public research is organised, the incentives of individual researchers, research specialisations and evaluation practices.

Evidence on whether agriculture research priorities are aligned to societal needs, and how they are determined, will become even more relevant in the future as international public agriculture research shrinks (Evenson and Gollin, 2003) and shifts from high to middle income countries, reducing more than proportionally in low income countries (Pardey et al., 2016);[5] while more than 800 million people in the world go hungry. In addition,

[4]Although attempts are made to elicit farmers' preferences when defining research priorities (e.g. Pingali et al., 2001).

[5]This reduction is in terms, also, of net overseas development assistance in agriculture (data from Bill

the number of marginal farmers, landless poor and urban poor (IFPRI, 2013) will increase and projected increases in yields will be insufficient to address the growing need for food (Ray et al., 2013). Finally, climate change will have the strongest effect in areas that host most of the world's malnourished population, depleting water resources and soil fertility, areas where the world poorest population mainly benefits from agricultural income (Ligon and Sadoulet, 2011). Pardey et al. (2016, p. 303) notes that: "One of the major global challenges in the years ahead is getting the relevant agricultural innovations into the hands of the world's poor farmers, such as those in south Asia and sub-Saharan".

In order to reshape science and technology research landscapes to contribute to technological trajectories that better address societal needs, we require a better understanding of the factors and actors that shape research trajectories. The present study is a first step towards providing statistical, high level and coarse-grained evidence on the relation between research priorities and societal demands.

3 Theoretical Framework

Following Sumberg et al. (2013), we can distinguish between two main perspectives on priority setting in agriculture. The first one, commonly employed by agricultural economists, assumes that priorities are the result of a rational process of selecting optimal outcomes given resource constraints. Several books and reports were published following the establishment of the Consultative Group on International Agricultural Research (CGIAR) to address questions about priorities and impact evaluation (e.g. Arndt et al., eds, 1977).

The second tends to emphasise how non-commensurable political and social forces (and power) shape research priorities (Thompson and Scoones, 2009; Frickel et al., 2009; Fuchs and Glaab, 2011). Research is seen as the outcome of interactions among several actors that influence priority setting. Even when policy makers are able to set agendas to maximise societal impact, their priorities may not be aligned to the preferences of agricultural scientists. Researchers appear to respond mainly to incentives from within the scientific community and peer pressure (such as their contribution and productivity, preference for fundamental research, prestigious publication outlets and notions of research excellence) (Raitzer and Norton, 2009).

From this second perspective, a number of scholars have studied agricultural research as the outcome of a systemic process, which depends on the interactions among actors, technological regimes that restrict research and technology choices and institutions (Sumberg, 2002; Klerkx and Leeuwis, 2008b; Vanloqueren and Baret, 2008, 2009; Klerkx et al., 2012, e.g) that shape both interactions and regimes (Dosi, 1982; Possas et al., 1996; Parayil, 2003).

While we acknowledge the relevance of power and systemic factors for shaping the direction of research, in a first approximation and for reasons that we hope will become

and Melinda Gates Foundation).

clearer when we discuss the empirical strategy, in this paper we make simplifying assumptions about investment decisions in national agriculture research.

Following Norton et al. (1992), we make the following three plausible assumptions. We assume that each actor in the Agriculture Research System (ARS) faces trade-offs when setting priorities, and obtains different returns from its research results.[6] Priorities can range from efficiency and economic growth, to the distribution of gains and several other societal demands and challenges (e.g., environmental protection, human nutrition, etc.). Some of these objectives may be complementary, but they are often substitutes: achieving one objective (e.g., aligning rice cultivation practices and output to international trade standards) may come at the expense of some other objective (e.g., reducing the costs of rice cultivation). Let us assume, also, that each actor in the ARS assigns a different weight to each objective (and may change these weights over time). Based on the trade-offs between objectives and the weights assigned to them, each actor defines its preferred allocation of resources in agriculture research. Finally, we assume that each actor has a different influence on the final funding allocation decision, which also can be measured with a weight. Then, the resulting funding allocation portfolio is a combination of the actors' individual weights (their individual portfolio choices) and the distribution of weights across actors on the aggregate allocation of funds (de Janvry, 1978). This framework assumes, also, that the actors involved in the allocation of resources can ascertain objectives in a straightforward way. As in the induced innovation model (Ruttan, 1997), policy makers and researchers are assumed, also, to know the preferences of the users of new research, which might be farmers or food consumers, that is, they know their demand for research and new technologies. However, as we suggest later, this assumption is not very relevant to our framework. What does matter is that different actors tilt the portfolio of agricultural research funding in one or another direction, however imperfect or misguided their information and choices.

The above description of decision making about the allocation of resources to research is suited to a world without frictions or sunk costs, and with free access to knowledge and research possibilities. In the real world, technological regimes (Dosi, 1982), bounded rationality (Arthur, 1994), technical bottlenecks to development and adoption of technologies (Binswanger, 1977), path dependency (Arthur, 1989) and continuous feedback mechanisms in the innovation process (Kline and Rosenberg, 1986) make it difficult to predict the allocation of resources to different research topics based on the preferences of a few policy makers and scientists. However, the crucial implication of the above framework is that the allocation of research funding reflects the combination of the priorities of the actors in an ARS. Due to frictions, sunk costs, tacit knowledge, path dependency and the like, we are not able to discuss how different actors and interactions influence the allocation of resources. Although there is some fascinating research on this very interesting topic (some

[6]de Janvry (1977) proposes a payoff matrix that represents the return from research to different social groups. The composition of groups, their representation in the political and administrative structure, and their influence on the political and administrative structure, affect the research direction.

of which was discussed in earlier sections), it is beyond the scope of the present study. Our approach is simpler and data driven, and takes the outcome of this complex decision process at face value, as an indicator of what this complex web of actors and economic and socio-technical constraints prioritised. We discuss measurement of the research landscape, emerging from this complex web, in Section 4.2 on data strategy.

So far, we have proposed that the allocation of resources reflects a combination of the priorities of different actors in the ARS. Next, we discuss how these priorities are formed. The literature discussed in Section 2 suggests that priorities can be guided by political factors (Sarewitz, 2010), the scientific community's agenda (Bush, 1945; Bozeman and Sarewitz, 2005), industry, research funders and public health organisations' agendas (Vanloqueren and Baret, 2009; Wallace and Rafols, 2016), researchers' incentives (Raitzer and Norton, 2009), foreign investors (Shiva and Crompton, 1998) and collaborations with international scientists (Evenson and Gollin, 2003; Brooks, 2011b; Herdt, 2012). Priorities can crystallise, also, around incumbent research routines and models, which may be sticky and slow to adapt to changes in the external environment Cowan and Gunby (1996); Vanloqueren and Baret (2008), or may be driven by serendipity Roberts (1989); Yaqub (2018). These are mainly supply side factors. In addition, we know from the literature on innovation that demand plays a relevant part in shaping research and innovation activities (Schmookler, 1966) and that most successful innovations are driven by both supply and demand factors (Mowery and Rosenberg, 1979). So to what extent does final demand (and which demand) influence the priorities of different actors in the ARS? Is it demand from wealthier farmers that export their crops, or demand from the poorest farmers that obtain low yields and experience adverse climatic conditions? Or is it demand from consumers who need access to more nutritious and cheaper food?

As discussed above, each actor in the ARS, when setting priorities, is likely to be influenced by different supply side factors and different demands. In this paper we are interested, especially, in the extent to which demands influence priority setting resulting from combining the decisions of all the different actors in the ARS.

Building on the simple and intuitive framework developed by agricultural economists (de Janvry, 1977; Norton et al., 1992), and acknowledging the complexities of priority setting in the real world, we propose the following empirical strategy. We use aggregate data that capture the final outcomes of rice research agenda setting, that is, research results measured by publications. We assume that publications capture reliably the outputs of research (evidence on this assumption discussed in the next section). Because publication rates can differ, *cæteris paribus*, across countries and disciplines or research areas, the results may suffer from some bias; we control for these persistent differences by including fixed effects in the estimations. We use publications for two main reasons. First, to our knowledge, there are no other data that allow exploration of research topics across countries. Second, publication data can be allocated to different administrative units, revealing information on global research, in regions, countries or sub-national systems. Here, we focus on the country level. By capturing the output of research, we reduce the complexity of priority setting, but also account for its result. Publications capture "revealed priorities"; we remain agnostic about the process that lead to them.

To study the relation between research priorities and societal demands related to rice, we aggregate country level composition of demand, ignoring its distribution across the society. We proxy actual demand for rice by shares of use for different purposes and a number of rice related farmer and consumer needs. We interpret these as "revealed demands".

We leave it to further research for more in depth investigation of specific countries' research pathways and how these are shaped by different political, technological and systemic forces (see Vanloqueren and Baret (2009) on prioritising genetic engineering over agroecological innovations for a brilliant example).

4 Data and Methods

4.1 Data

The identification of rice research output topics is based on publications recorded in the CAB Abstract database. For several reasons, we use CAB, rather than the more commonly-used Web of Science and Scopus. First, CAB focuses on agriculture and environment, which has significant advantages for coverage and classification compared to the standard databases. Second, CAB contains a specific collection of publications, classified as explicitly about rice (subject code 7U). This allows a cleaner dataset that excludes a significant number of publications containing words related to rice – rice or *oryza* – but which are not focused on rice. The records are selected and assigned by subject specialists. Third, CAB translates into English a consistent number of abstracts from other languages, which substantially increases coverage of non-English speaking countries' published research. CAB covers both well-known and independent and learned publishers, including a consistent number of outputs that do not feature in standard journals that include agricultural research, and which can go unnoticed in international scholarship, but which provide a better understanding of the local focus of research. Fourth, as a result, although not perfect, CAB offers a better understanding of local research than either Web of Science or Scopus which cover significantly lower numbers of publications and in which less developed countries are under-represented (Rafols et al., 2015). Fifth, CAB librarians manually classify all papers using descriptors from a given thesaurus (subject indexing) and broad subject codes, which increases the precision of the co-word analysis employed to derive topics (more below). Finally, to our knowledge this is the first large scientometric study to use the CAB Abstracts database.

After cleaning the publications data of duplications and entries with missing information, and limiting the analysis to 1983-2012 – to avoid biases due to coverage – we are left with 105,356 documents that discuss the results of research on rice.

To identify societal demands, we use detailed data on agriculture, particularly on rice, from FAOSTAT (on production, inputs, and food balance).[7] This includes information on land use, yield, production, balance sheets (uses of rice including final consumption, trade, seeds, etc.), the use of variable inputs and capital, R&D, prices and nutrition.

Finally, we also use indicators form the World Development Indicators (WDI) such as education expenditure and agriculture value added.

4.2 Methods, Definitions, and Assumptions

The analysis is conducted in two main steps. First, using the full set of publications (for all countries and years) we identify the main topics of rice research across the globe between 1983 and 2012. Second, we exploit variation in the distribution of research outputs over the identified topics, across countries and years, to study the extent to which countries' revealed research priorities are related to revealed demands for improvements in the rice technology. Below, we discuss the identification of research topics, revealed research priorities and demands, and the assumptions behind these definitions.

Identification of Topics in Rice Research

We define a research topic as a *broad area of research which shares several aspects of the (rice) technology, the issues addressed by the technology and/or the method of enquiry.* Technology aspects might be the seed or the soil properties, for example. The issues addressed could be improvement in yields or increased nutritional properties, for example. The method of enquiry is related to the discipline – for example, biology or economics – although in some cases the output may be multidisciplinary.

Due to data availability, we focus only on published research, whether in a journal, conference proceedings, a bulletin or other miscellaneous output. This excludes grey literature which is not included in outlets covered by CAB. However, CAB remains the best option to cover some grey literature, and especially publications from the Global South: in the context of rice, Rafols et al. (2015) document that less than 85% of publications in CABI are journal articles, while in the case of Scopus this is nearly 94%. Better coverage and understanding of the grey literature is needed, but would require intensive data collection within countries and is beyond the scope of this research.[8]

Therefore, we assume that publications – in the broad selection of outlets covered by CAB – are an accurate representation of the research on rice globally, and in individual countries. Because publications do not cover all the research performed in an economy, our paper focuses on the ARS within the Agriculture Innovation System (AIS) (Vanloqueren and Baret, 2009) and excludes several, less formal, research actors (including farmers). Within the ARS, our paper provides results on a selected number of actors more likely

[7] Data available from the FAOSTAT website: http://www.fao.org/faostat/en/#data, accessed February 2016.

[8] Google Scholar is a relevant option to capture more grey literature, but black-boxing regarding its coverage and inclusion of different languages, and problems related to selecting relevant results based, e.g., on the search word "rice", militates against its use in this study. We hope to extend the analysis to Google Scholar in future research.

to publish on rice, such as university and public and private sector research organisation researchers. Thus, we exclude ARS actors less likely to publish, such as private companies. The use of publications to investigate research trajectories is not new (Mina et al., 2007; Rotolo et al., 2017; Cassi et al., 2017). Similar to how patents are used to investigate the technology trajectories (Graff, 2003; Verspagen, 2007; Fontana et al., 2009), publications are considered the best proxy to study the direction of scientific investigation in given research fields. We make the small additional assumption that publications provide an indication of the investment made in a given field of research (see discussion below on the relation to research funding). Earlier research has used publications to document differences in the way scientists prioritise across different topics of research (Arvanitis and Chatelin, 1988).

In order to identify topics and their relations we use a co-word algorithm to analyse the similarity among documents. Co-word analysis is used frequently to map and study similarities between outputs in a given scientific knowledge domain, using the keywords employed in the outputs in which the results of research are published (Callon et al., 1983) – and which are assumed to refer to the scientific content of the research.

In a nutshell, two descriptors are considered similar if they appear together in several documents. For example, a paper on `genetic sequencing`, `varietal resistance` and `genetic mapping`, and a paper on `genetic sequencing`, `varietal resistance`, `genetic mapping` and `pest control` are similar and could be on the same topic. If `genetic sequencing`, `varietal resistance` and `genetic mapping` appear together in a large number of papers, they likely form a topic. If `pest control` appears in a subset of papers, research on pest resistance is probably related to molecular biology, but these are two different topics. Based on similarity, a co-word algorithm clusters terms into topics, similar to cluster analysis.

The choice to use descriptors with respect to titles and/or abstracts, improves precision in identifying topics.[9] CAB librarians add descriptors to each document, choosing from a closed and controlled set of terms (thesaurus). On the one hand, descriptors define a publication topic quite precisely, capturing aspects of the technology that may not be included in short titles or abstracts. On the other hand, the use of descriptors eliminates a large amount of the noise usually present in text analysis, due to the presence of terms that may not be relevant to the topic (Leydesdorff and Hellsten, 2006).

In our corpus of 105,356 documents on rice, there are over 10 thousand descriptors. To improve clustering precision, we drop descriptors that are not useful for discriminating across topics because they are too frequent (e.g. rice or cereal); we also drop descriptors that are too infrequent and appear in an insufficient number of documents, and therefore do not inform about the similarity between two publications. There are several ways to optimise the number of terms (in our case, descriptors) in semantic analysis (van Eck et al., 2010b,a). We choose an approach which is similar to the one implemented in VOSviewer

[9] We tried clustering using titles and abstracts; this identified the same topics, but the borders between topics were more blurred.

(van Eck and Waltman, 2010), the software we use for clustering and mapping topics. We adapt it to the specificities of our corpus of publications from CAB (a repository not covered by any of the known software and packages for semantic analysis or scientometrics more generally), the information used (descriptors rather than terms) and the analysis.[10]

First, whereas VOSviewer extracts relevant terms from the document's fields, we employ well-defined terms, the descriptors assigned by expert CAB librarians: we do not need to make assumptions about the relevant terms in a published document. Second, because we want to investigate how the focus on topics changed through time, we need also to consider that the relevance of terms may have changed through time. For instance, analysing the entire database (83-12), it might be that the term `water` is not too frequent across papers, therefore not irrelevant. However, the distribution of the term water may be skewed across years: in a given period, it might be too frequent to discriminate among publications, while in another given period, it might be too infrequent to add information. Were we to cluster terms according to their relevance throughout the whole period, we would keep `water` in the analysis. In the sub-period when research on `water` is very frequent, it might appear in almost every paper, so would not be related to any other particular term (but rather to all of them); whereas in the period with infrequent reference to `water`, it would provide too little information on its similarity with other terms. In neither period would the term be relevant. Even more crucial, the term `viscosity`, for example, may be too infrequent in the entire database, but may be a crucial topic in one or two specific periods to discriminate among research on different technologies. Were we to cluster topics from the entire period, we might exclude a term that is relevant.

Taking account of both these issues, we divided the data into sub-periods with similar numbers of terms.[11] For each period, we created a co-occurrence matrix of descriptors that occur at least five times in that period, using the R package 'mpa'.[12] The *co-occurrence* matrix is a matrix in which all descriptors (occurring more than 5 times) appear in both columns and rows; each cell reports the number of documents in which a pair of descriptors co-occurs in the same document; the diagonal reports the *frequency* of each descriptor, that is, the number of documents in which it appears (across all documents).

Using frequency and co-occurrence of each descriptor with other descriptors, we compute the relevance of a descriptor in a given period measured by Term Frequency-Inverse Document Frequency (TF-IDF) (Salton et al., 1983). TF-IDF is an efficient method to identify terms that, in one sub-period, co-occur frequently with other terms, but do not occur in all documents. On the one hand, a term that occurs in too many documents (e.g., rice), has a very similar probability to appear with any other term and does not

[10]We are grateful to Ludo Waltman for support and suggestions.

[11]These are 1983-87, 1988-91, 1992-95, 1996-99, 2000-02, 2003-04, 2005-06, 2007-08, 2009-10 and 2011-12. The difference in period spans is due to the increasing number of publications in CAB and, therefore, the increasing number of terms per year. In order to reduce the bias due to different numbers of terms in different periods, we opted for a similar number of terms per period, allowing for some differences in time span.

[12]This allows to produce a matrix with a given set of terms, without the need first to extract the terms from the text. Documentation is available at https://cran.r-project.org/web/packages/mpa/mpa.pdf

add information about which terms more frequently occur together. On the other hand, a term that appears in a limited number of documents, in each document paired with few other terms, contains relevant information on the probability that two terms appear together and, therefore, are likely to be used in research on the same topic. In sum, relevant descriptors are loaded with information to define topics.

We compute TF-IDF for each term k in period t as TF-IDF$_{kt}$ $= \ln(TF_{kt}) * IDF_{kt}$. Where TF_{kt} is the frequency of term k in period t, that is, the number of documents in which k appears in period t;[13] $IDF_{kt} = \ln(M_t - 1/TF_{kt})$, where M_t is the total number of terms in period t (that appear in more than 5 documents). The basic idea is that, if a term k occurs in many documents with many other terms (high TF_{kt}), this means it might be quite relevant to research on rice, but might also be because it appears in documents that are more general and therefore include many descriptors. Therefore, TF-IDF$_{kt}$ increases due to an increase in TF_{kt} (a relevant term), but decreases as TF_{kt} converges to M_t. When a term k appears with all other terms, $IDF_{kt} = 0$.

Finally, in each period we define as relevant only terms with TF-IDF$_{kt}$ above the median.

Next, we generate the co-occurrence matrix for the period 1983-2012, selecting only relevant terms (based on their relevance in the sub-periods). From the co-occurrence matrix, we cluster terms into topics using the VosViewer procedure (detailed in Waltman et al. (2010)) and map topics and their frequency and relation. Each cluster defines a rice research topic.

Identifying Countries' Research Priorities

The next step consists of measuring the relevance of each topic in each country/year. Do all countries dedicate the same resources to research on the same topic? Does the rice research landscape of research change through time, within and across countries?

We assume that research output (publications) is correlated closely to research inputs. If a country's ARS organisations invest x per cent of their funds in investigating topic j, we expect to observe a share of publications in topic j similar to x. Employing a strategy, commonly used in economics to identify unobservable consumer preferences on the basis of their observable choices (Samuelson, 1938), we refer to a country's i observable share of publications in topic j in year t as the country's *revealed* (non-observable) relative priority for that topic of research on rice. Therefore, the distribution of publications per topic in a given country/year is interpreted in this paper as the distribution of revealed priorities in rice research for that country/year.

Unfortunately, we do not have access to detailed funding data that would allow us to verify this assumption for rice – the reason for using publications is to proxy for the priorities, which, ideally, we would like to observe directly in research inputs (funding), but on which data are limited. Using recent data collected by Digital Science and accessi-

[13] One term cannot appear more than once in one document, by definition.

ble via the Dimensions service, we investigate the relation between research funding and publications output in the US (for which we have reliable data). Figure 3 in the Appendix plots the relation between the number of grants (panel 3a) and total funding (panel 3b), and the number of publications, across Health Research Classification System (HRCS) classes, between 2014-17. Although in-depth analysis of this relation is beyond the scope of this paper, Figure 3 shows a rather salient linear relation between inputs, in terms of funding (National Institutes of Health – NIH priorities) and publications outputs. Disease classes that receive more funding are more likely to publish more, and the relation seems linear and robust – although a deeper investigation is required. NIH is a perfect study case, because both funding and publications are categorised using the HRCS, thus, the results are quite precise.

Appendix Figure 4 depicts the same relation for agriculture. In panel 4a we plot the relation between number of grants in the US (the country with most information on funding data, and one of the only two countries (with Japan) that reports information on funding for rice) for a given crop and the number of publication on the same crop, in the same period (2013-17). In panel 4b we plot the relation between US$ millions of research funding in each crop, and the (log) number of publications on the same crop, in the same period (2013-17). Confidence intervals are slightly larger than in the case of NIH, and the linear fit is slightly less strong, but Figure 4 suggests that there is a positive and linear relation between the amount of funding allocated to a crop and the resulting publications.

In sum, through analysis of linkages between research grants and publications provided by Dimensions, using NIH data from Federal Reporter (also known as StarMetrics), we can confirm a linear relation between amount of funding and the number of publications by disease class. We would expect that, in the absence of information on the funding agency (when pooling all the agencies), as in the case of agriculture funding for different crops, a similar linear relation might be less likely. Instead, we find a linear relation also between US funding and publication shares across crops. Also, data from the pharmaceutical sector (Schuhmacher et al., 2016) show a correlation between R&D expenditure and number of new molecular entities. It is possible that, at a more fine grained level (fields of research within a crop), the relation might weaken. We hope more fine grained data on research funding will become available for research, which would allow us to study the relation between research topics and societal needs without relying on research output proxied by publications. However, publications (from non-biased repositories) currently are, we believe, the best proxy to reveal the research system priorities.

To compute the distribution of publications per topic in a given country/year, we first allocate publications to different topics, proportional to the number of relevant terms in a publication that were clustered in that topic. If a publication h contains N terms k, its contribution to a topic j is weighted by the number of terms clustered in j: $\sum_{k \in j} k/N$. Then, we measure the share of topic j in period t as the proportion of publications in the topic, in the total number of publications in the same period. Next, we allocate publications to countries, using the affiliation of the first author. Unfortunately, we cannot (as we do for topics) use fractional counting for countries (i.e., allocating one publication to multiple countries, proportional to the share of authors) because one of the main limitations of CAB data is that they provide information only on the first author's affiliation. We believe that the advantages in terms of country coverage, especially rice producing developing countries (as documented in Rafols et al. (2015)), more than compensates for the relatively uninfluential missing information on co-author affiliations. In particular, because middle and low income rice producing countries represent the majority of the observations in our econometric analysis, the severe under-representation of publications from these countries is likely to bias the results badly and substantially more than the small changes attributed to the fractional allocation of publications to all authors' affiliations. See Appendix C for a discussion why this is unlikely to affect our results.

To measure a country's revealed research priority in a topic with respect to other topics, we compute the relevance of one topic in a country/year as an index of relative specialisation (RSI_{jt}) using the same formula as for the revealed comparative advantage index (Balassa, 1965): $RSI_{jt} = \frac{P_{ij}/\sum_j P_{ij}}{P_j/\sum_{ij} P_{ij}}$, where P is the number of publications in country i in topic j.[14] Given the fluctuations in publication timings, especially for smaller countries, RSI_{jt} is computed as a three-year compounded average. If a country i has a relative specialisation in topic j it means that, with respect to overall specialisation, i publishes more on topic j. The larger the RSI_{jt}, the greater the country's specialisation in topic j. Changes in RSI_{jt} for all topics indicate changes in the country's revealed research priorities for rice technology. RSI_{jt} then indicates a specific instance of the revealed research priority in time period t.

Identification of Countries' Research Demands

Similar to how an individual's revealed preferences for different goods can be deduced from individuals' purchasing behaviour, we can deduce a country's revealed demand for advances in a specific aspect of rice technology from that country's relative use of total rice production (composition of aggregate demand) and a number of features of rice cultivation and human consumption that indicate the country's relative needs with respect to rice. We acknowledge that different farmers and consumers in a country have different priorities, but our level of analysis in this paper is the country, not the individual – the relation between a country's ARS priorities and a country's demands. We then need to aggregate these individual demands. Using available data, we do this by taking either the country average (rice use) or the country total (relative needs). We cannot control for the fact that some demands may come from specific areas in the a country, which may be particularly neglected by the country's ARS because of their smaller representation and greater marginalisation. However, we hope to do this in future work.

Ours is not the first paper to refer to societal (or human) needs, at the national or

[14] As a robustness check, we also produced estimates using the relative percentage of publications in year t in topic j.

global level, and their relation to agricultural research. Indeed, a large part of research in agriculture is shaped by responding to the population's needs (Herdt, 2012) and, in many case, to those of national or subnational components: for example, consumer and producer needs (Welch and Graham, 1999), hunger and poverty (IFPRI, 2013; Fischer and Hajdu, 2015), food (Herring, 2007; Ray et al., 2013; Carlisle and Miles, 2016), nutrition (Römheld and Kirkby, 2010), soil fertility (Fuller et al., 2010), yields and productivity (Fuller et al., 2010; Kolady et al., 2012), income (Herring, 2007), environment (Herring, 2007) and innovation (Klerkx and Leeuwis, 2008a). There is a stream of literature criticising the approach used to address agriculture related problems as the needs of the whole population; for instance, Brooks (2011a) discusses this problem with reference to malnutrition. Similarly, in the health literature, scholars tend to talk about the focus of research in relation to health needs across and within countries Gillum et al. (2011); Evans et al. (2014); Cassi et al. (2017). If at least part of the research conducted on agriculture since the green revolution were to be framed in terms of farmers' and consumers' pressing needs, we would assume that country level information on how the rice is used and on rice related farmers' and consumers' needs, might reflect existing demand. In this paper, we test whether this relation between ARS priorities and societal demands holds in the case of rice.

Using data from FAOSTAT, we distinguish between *relative use/demand* of a country's rice production: *export*, *food* (direct intake), *seeds*, *waste*, *imports*, *processed food* and *animal feed*.[15]

Table 1 summarises the variables used (col (a)), the observed measure (col. (b)), societal demand for which we assume these variables proxy for (col. (c)), and the research topic that we expect will be the country priority if the observed variable is higher (H) or lower (L) (col. (d)). The topic names are based on the results of the co-word analysis, which is discussed in the next section (5.1). For example, if most of the rice in a given country is exported, this might suggest that national demand for nutrients from rice is not particularly high. However, if most of the rice produced is consumed domestically and imports of rice also are high, this might indicate that demand from local population for nutrition from rice is high. In the first case, we would expect human consumption not to be a priority for the national ARS; in the second case, we would expect research on human consumption to be a priority – to feed the local population. Similarly, relatively higher use of rice for seeds (with respect to other countries), following the idea of revealed preferences, would suggest that, to some extent, there is a relatively higher demand for seed technologies or for access to seed. In terms of priorities, we might expect the ARS to invest in traditional genetics or genomics to improve seed quality and reduce their price.

Using FAOSTAT and World Bank (WB) WDI data, we created additional indicators of revealed demand for rice technology: *Rice calories* is a measure of per capita human intake of daily calories from rice. The higher this value the higher the relevance of rice for

[15]Due to data availability, the analysis focuses mainly on the first five production categories: *export*, *food* (direct intake), *seeds*, *waste*, *imports*, with *processed food* and *animal feed* estimated separately. The final category, *Other*, is not considered here.

the nutrition of the average individual. Although measures of protein or vitamin intake would have been more accurate,[16] caloric intake is a good proxy for the relative relevance of rice in the human diet and, therefore, of how improvements in the rice technology might improve human nutrition. In countries where rice calories are higher, we assume that use of rice for food is also higher, suggesting demand for improvements to the nutritional features of rice. *Rice yields* is a measure of rice productivity: although it may depend on several aspects of the rice technology (including how it is cultivated and in which ecosystem), low rice yields suggest there may be a demand for improved yields, irrespective of geographical and climatic conditions.[17] Because this is a country average, if the distribution is very skewed, demand may come from those farmers who experience extremely low yields, but not those that enjoy high yields. Unfortunately, we cannot control for these differences, but it seems reasonable that a country facing low average rice yields might need to prioritise this area in its research portfolio, relatively more than a country with high average yields. *Fertilisers* is the total amount of chemical fertilisers used per arable land and permanent crop areas.[18] – Unless we assume that fertiliser use is imposed, as might be the case in planned economies, higher use of fertiliser suggests higher demand from farmers, which, in turn, might suggest relatively higher demand for research on plant nutrition. *Rice area* is the per capita amount of arable land devoted to rice, suggesting the overall relevance of the crop in the country, as source of income.

A number of indicators are estimated separately because they are available only for a limited number of countries. *Under-nourished* measures the percentage of undernourished population, which we think suggests demand for improving the nutritional content of crops. *Pesticides* is the total amount of chemical pesticides used per hectare of agricultural land.[19] Similar to the case of fertilisers, higher pesticide use suggests higher demand for them from farmers. In turn, this might suggest higher demand for research on plant protection.

The reliability of FAO data depends on the quality of the data reported by countries (Hannerz and Lotsch, 2008), which suggests some caution when interpreting the results. However, our analysis refers to national level gross figures and does not aim to assess local agro-economic conditions in detail. Also, as discussed below, the country figures are quite stable and fairly close to expectations about uses and potential farmers' and consumers' needs for rice. More micro level data is needed for a more precise identification of the relations between research priorities and societal demands.

[TABLE 1 HERE]

[16]Unfortunately, the data available cover too few countries/periods.

[17]Note that our objective in this paper is not to evaluate the impact of research, but whether it is related to the problems faced by farmers and users. Low productivity would be a problem that concerns both and would increase demand for research on productivity. How this research is implemented will depend on the specific research programme, which is not relevant to our study.

[18]We use national aggregate figures since this figure is not available only for rice.

[19]Again, we use national aggregate figures because this figure is not available only for rice.

Relation between Countries' Research Priorities and Demands

Having mapped rice research, globally and across countries, and estimated a number of potential needs, we next investigate the relation between countries' revealed research priorities and societal demands. We use the equation below to estimate the relation between revealed demands and revealed priorities (as defined above) across the 17 main rice producers and publishers[20][21] over 28 years.[22]

$$RSI_{jit} = \alpha + \Delta D_{jit} + \Xi X_{jit} + \pi_i + \tau_t + \epsilon_{jit} \quad (1)$$

where D_{jit} is a vector of revealed demands; X_{jit} is a vector of the country level control variables; Δ and Ξ are vectors of the coefficients; τ are time fixed effects to control for global trends in technologies related to rice, changes in publication practices or any other unobservable change common across countries; π are country fixed effects to control for unobservable country specificities, such as rice cultivation techniques, national publication practices, etc., not captured by the controls; ϵ is a country specific error term; and α is a constant. All variables are in logarithms.

We estimate the above equation using a panel fixed effects estimator. Research specialisation in one period is likely to build on previous specialisation, which may cause autocorrelation. Also, research in one country is likely to be influenced by the research on the same topic in other countries, causing cross sectional dependence. Therefore, we correct standard errors for heteroscedasticity, autocorrelation and cross sectional dependence using the method suggested by Driscoll and Kraay (1998).

There may also be a time lag between manifestation of societal demands, their capture within the ARS, and the resulting research and publications. However, the level of aggregation (the country) at which we consider societal demands, is such that we do not expect them to change dramatically, but only slowly. Were we able to capture local and more disaggregated shocks to societal demands, such as, for example, droughts, floods or pest infestations, lagged variables may be more relevant. We checked visually for the pattern of changes in our proxies for societal demands.

Figure 5 in the Appendix plots the dynamics of the two main uses of rice across countries: food consumption and export. As expected, the series are relatively sticky, although some countries experience change in use along the 30 years. For instance, exports from Egypt, India and, later, Brazil have increased substantially, reducing the share of rice used for domestic consumption.

The patterns for less relevant categories of use (in relation to total production) such as seeds (6a), waste (6b), import (6c) and animal feed (6d) are also sticky; the larger variations observed are due mainly to the reduced scale (Figure 6 in Appendix). For instance, Brazil experienced fluctuations in the percentage of wasted rice, and Nigeria's

[20] According to CAB and FAOSTAT data, the main rice producers are often, but not always also the main contributors to global publications.

[21] These are: Bangladesh, Brazil, China, Cuba, Egypt, India, Indonesia, Iran, Japan, Nigeria, Pakistan, Philippines, South Korea, Taiwan, Thailand, USA, and Vietnam.

[22] We lose 1983 and 2012 as a result of computing three-year averages.

rice imports followed long cycles. Overall, the allocation of rice to different uses changes over time, but (with a few exceptions), relatively slowly. This suggests that countries' research priorities may also change quite slowly (and may follow relatively predictable trends).

The variables proxying for revealed demands related to farmers' and consumer's potential needs, are likely to be even more stable and predictable through time. Diets tend to change slowly, especially when they are culturally embedded. Yields tend to improve through time, but slowly as farmers change their technologies and cultivation practices. Appendix Figure 7 plots changes in per capita consumption of calories from rice (in logs) (7a), amounts of chemical fertilisers per arable land (in logs) (7b), rice yields (7c) and the total arable area dedicated to rice (in logs) (7d). In most countries, the number of calories consumed from rice remain quite stable, although they differ substantially across countries. Relative exceptions are the USA, Nigeria and Egypt, which show increased importance of rice in the national diet. As expected, yields tend to grow in most countries, at a different pace, but following a relatively stable trajectory. As a result, the area used for rice reduces, at a similar rate. Chemical fertiliser use is less stable and shows a decrease in the late 1990s for all countries.

In sum, the revealed demands seem to suggest that were research priorities aligned to them, they would evolve through time, following relatively stable country level demands. In other words, aggregate societal demands seem persistent and are expected to influence the research agenda in each time period. However, we checked also for a number of lags in the estimations (between 1 and 5); the results are discussed below.

Given the high level of data aggregation, in this paper we provide simple correlations and leave investigation of causal relations between societal demands and research priorities to further analysis.

5 Results

5.1 The Global Map of Rice Research

Using a 30-year time window (1983-2012), we use CAB descriptors to create a global map of rice research, showing the relative size of the main rice research topics world-wide and how they are related. Figure 1 shows the six major topics identified by the text analysis: (i) plant protection, divided into pests and weeds (north on the map) (*plant protection*); (ii) production practices and socio economic issues (south-west on the map) (*production practices*); (iii) plant nutrition and yields (north-west on the map) (*yields*); (iv) rice varieties and traditional genetics (north-east on the map) (*traditional genetics*); (v) transgenic, molecular biology and genomics (south-east on the map) (*genomics*); and (vi) consumption, human nutrition and food (south on the map) (*human consumption*).

[FIGURE 1 HERE]

As might be expected, plant protection is closely related to traditional genetics research. However, the connection to the more recent genomics is looser. A large part of genomics research seems to be only weakly connected to most the rest of the map, apart from traditional genetics. In the west of the map, plant nutrition, yields and production practices are closely linked. Some of the research in these topics overlaps with research on human consumption, which is relatively disconnected with the remaining research topics.

In terms of the relative importance of the different topics, yields and genomics are the main topics world-wide and across the 30 years analysed. Figure 8 in the Appendix plots the density map, which provides a clearer visualisation of the relative frequency of publications in each topic. Traditional genetics and plant protection are ranked next, followed by human consumption and production practices.

This distribution of publications across topics captures global research priorities over the three decades examined, but is likely to hide crucial differences among countries and over time. Have global priorities changed through time? To be sure, genomics research on rice is more recent than traditional genetics research. Do countries have similar research priorities? Do country priorities change over time? If so are these changes correlated?

5.1.1 Country Variations

We focus on a selection of the countries that are the main rice producers, or publishers (both in most cases).[23] Figure 2 plots the share of publications per topic over time for the 17 selected countries (and the average across countries).[24]

[FIGURE 2 HERE]

Some topics (plant protection, to some extent production practices and, most recently, traditional genetics) become less relevant globally; others (such as genomics and human consumption) increase across countries; and others (such as plant nutrition and, until recently, traditional genetics) remain quite stable. Beyond these average tendencies, there are large differences in revealed research priorities among countries.

We exploit these differences over time and across countries to study how the focus of published research (revealed research priorities) is related to the country's main revealed demands for rice technology.

5.2 Relation between Research Topics and Countries' Demands

In this section, we discuss results of the estimations of the relation between countries' revealed research priorities and demands, as defined in Section 4.2. We exploit the variation in revealed research priorities and demands across countries and over time. We estimate our main equation for each of the six topics identified, in order to show whether revealed

[23] Figure 9 in Appendix A.2 plots the evolution of country rankings with respect to publication numbers, from 1983 to 2012, and the total number of publications per year.

[24] To improve visualisation and reduce the relevance of short-term fluctuations in publications, each series was estimated using Locally Smoothed Regression (LOESS).

demands are related differently to the different prioritisation of research/publications topics.

Table 2 reports estimates of the relation between revealed demands, measured as relative use of total rice research output[25] and the Relative Specialisation Index (RSI) of any of the six topics.[26] The results are for the 11 countries for which rice use data are available, over 28 years.

[TABLE 2 HERE]

When we control for education, value added in agriculture (strongly correlated to per capita income), percentage of agriculture value added in the economy, and average cereal yields, we find a limited relation between relative use of rice and country specialisation in different topics. This could be expected for plant protection (col 1) and yields (col 2) since these topics are not related directly to a specific use of rice output, but is more surprising in the case of the other four research topics.

In particular, we found no regularity among countries that are specialised in genomics (col 4): the focus on genomics seems unrelated to the main uses of rice, whether for export, consumption, seeds or some other. Not surprisingly, countries that specialise in traditional genetics (col 3) use a significantly higher share of rice output for saved seed than do other countries. However, again there is no significant pattern related to export, import or consumption.

The most interesting results are for human consumption (col 5) and production practices (col 6). In the former case, on the one hand we find no evidence that countries that use relatively more rice to feed their populations conduct more research on human consumption (if anything the reverse applies). Instead, countries specialised in human consumption research tend to use more rice for seeding purposes: it is perhaps the case that the research is related to increased nutritional aspects of new seed varieties. On the other hand, we find that countries that manage to reduce waste (improve storage technologies) have a stronger research focus on human consumption. We run a separate regression to analyse the relation between the percentage of rice used for food processing (agro-industry) and specialisation in human consumption research.[27]. On this subsample of countries we find a positive and significant coefficient, suggesting that research on human consumption (and food technologies) is focused mainly on agro-industry rice processing, not direct consumption.

Specialisation in published research on production practices (policies and the consequences of rice cultivation rather than rice technology) is significantly and positively related to lower use of rice for export and seed, and a higher use of rice for food. That is, countries that need their rice production to satisfy internal demand, where farmers tend to

[25] We distinguish between export, food, seeds, waste and imports.

[26] Estimations using the relative percentage of publications in a given topic, rather than the RSI, show very similar results (see Tables 5 and 6 in Appendix B.2).

[27] We run separate regressions because data on the share of rice used for food processing are available for a smaller number of countries. Results available from the authors.

import seeds, tend to focus their research on issues related to increasing the social impact of rice cultivation among both farmers and consumers.

The results for the relation between revealed demands proxied by relative needs satisfied by rice[28] and the RSI in one of the six topics, yield more interesting insights (Table).[29] The results are for the 16 countries for which data on revealed demands are available, over 28 years.

[TABLE 3 HERE]

Countries where we expect high demand of nutrients from rice (because per capital caloric intake from rice is higher than the average), specialise less in traditional genetics, genomics (cols 3 and 4) and human consumption (col 5). In other words, where there are nutritional demands for rice, this is not being addressed by research on molecular biology or food technologies. Instead, as previous table shows, higher demand for calories is addressed by focusing research on production practices and socio economic issues (col 6). Other areas of research that might improve productivity, such as plant protection (col 1) and yields (col 2), are also not significantly related to demand for calories from rice.

Due to data limitations, we ran separate regressions for the relation between the percentage of undernourished population and relative specialisation.[30] We found that malnutrition is positively and significantly related to specialisation in genomics (but negatively related to research on plant protection).

Countries seem more efficient at addressing rice yield than human nutrition. Those countries with lower than average yields specialise in yields plant nutrition (col 2), traditional genetics (col 3) and production practices (col 6).

Also, as expected, countries that use more than the average quantities of fertilisers (suggesting a need to increase yields), specialise in production practices (col 6). However, they seem not to prioritise yields (and plant nutrition) research more than other countries (col 2). Again, due to data limitations, we ran separate regressions for pesticide use. As in the case of fertilisers, it seems surprising that countries with higher than average use (per cultivated area) of chemical pesticide, suggesting higher demand for plant protection technologies, specialise significantly less than other countries in plant protection.

When we examine the overall relevance of rice, proxied by the relative area dedicated to rice cultivation, the results are less clear-cut (possibly because they are driven by China which focuses strongly on genomics and much less than average on plant protection).

How do the results change if we consider a lag between revealed demands and research priorities? We re-estimate Equation 1 using one to five lags. The results for the first and second lags are interesting, but become less reliable as we increase the number of lags.[31]

[28] We distinguish between calorie intake from rice, rice yields, relevance of rice cultivation measured by the relative area of rice cultivation, and use of fertilisers to nurture the plants.

[29] Estimations using the relative percentage of publications on a given topic rather than the RSI, show very similar results (see Tables 5 and 6 in Appendix B.2).

[30] Results available from the authors.

[31] Results available from the authors.

We discuss the main differences with respect to the baseline results, for revealed demands with one and two lags (Appendix Tables 7, 8, 9, and 10). Concerning revealed demands measured as the share of rice production for different uses (Tables 7 and 9 in Appendix), the differences are not significant: the negative relation between higher shares for food use and genomics, and import share and production practices, which were barely significant for D_{ijt}, are non-significant for D_{ijt-1} and D_{ijt-2}. Countries with higher import shares in both $t-1$ and $t-2$ have a smaller focus on plant protection, which is counter-intuitive, or may indicate that the reason for importing rice is unrelated to pests or weeds. Finally, countries with a higher use of rice for seed in $t-2$ do not have a stronger focus on production practices, which does not add much to the picture.

In terms of revealed demands measured as relative needs, the nuances are more interesting (Tables 8 and 10 in Appendix). First, the negative relation between high demand for nutrients from rice, and genomics (col 4), becomes non-significant for higher demand in both $t-1$ and $t-2$. The negative contemporaneous relation discussed above, may be due to the negative relation between income per capita and rice caloric intake (countries with more diversified diets, with respect to rice, tend to be richer), which is not completely captured by the control variables. Ultimately, specialisation in genomics is related only to the rice area size. In relation to the results for yields (countries with lower than average yields specialise in research related to yields (and plant nutrition0 (col 2), traditional genetics (col 3) and production practices (col 6) (Tab. 3)), countries with higher than average yields in $t-1$ and $t-2$ are significantly more likely to focus on plant protection (col 1). This result holds even with a five year lag, suggesting that the best performing countries may need to maintain high standards of plant protection to maintain high yields, more than any other research topic. Also, countries that use more fertiliser in $t-1$ and $t-2$ tend to focus on human consumption research, after one and two years, a result that is not particularly consistent with the alignment between societal demands and revealed priorities.

In summary, revealed demands for rice research, measured by relative use of rice and relative needs, only to a limited extent explain the country's revealed priorities in rice research, measured as relative specialisation of publications in a given topic. Table 4 summarises the results. We build on Table 1, where we summarised the variables used to proxy for revealed demands, and what topics we expected to be associated to each revealed demand. Table 4 reports the expected topic and the associated topic, distinguishing between positive and negative estimated associations between the research topic and the revealed demand.

[TABLE 4 HERE]

In some cases, the research specialisation is neatly aligned to revealed demands. For instance, countries that use large proportions of rice for food processing, specialise in human consumption (and food technologies) and countries that use large proportions of rice for food specialise in research on production practices, while countries with lower yields invest more than others in yields, traditional genetics and production practices.

However, we found also significant misalignments between countries' revealed demands and rice research priorities. The most prominent is the case of nutrition. Countries that use rice production mainly for food specialise only in production practices. In particular, they are not specialised in human consumption. Instead, countries where rice is an important source of caloric intake, specialise less in human consumption and, also, traditional genetics and genomics. Countries that use more fertiliser do not focus on yields, and countries that use more pesticides focus significantly less on plant protection.

Particularly relevant are the results on genomics – one of the topics with the largest number of publications in current research on rice technology. We found no relation between country specialisation in genomics and revealed demands for rice research. The only variable correlated with research specialisation in genomics is the size of rice cultivated area with respect to other crops, such as for China. Research on genomics seems not to be relevant for most of the revealed demands that we measured (except malnutrition for the smaller number of countries for which data are available).

6 Conclusions

This paper provides a first exploratory estimation of the relation between research priorities and societal demands in agriculture. We focus on rice in the period 1983 to 2012. We use information from a comprehensive publication database (CAB Abstracts), which is rarely used for scientometric analysis, despite its relative advantages for studying agricultural research, particularly in terms of coverage of non-English speaking and developing countries. We proceeded in three main steps.

First, we identified research priorities. Because we cannot observe the preferences of policy makers when setting priorities, or those of the researchers, we proxy the complex process of priority formation by research output, in this case publications. We refer to the distribution of publications over research topics as revealed research priorities, because it reflects the aggregated outcome of a variety of processes leading to research allocation across different topics. Priority setting depends on a complex set of actors and their interactions, ranging from policy makers, funders, researchers, research collaborations, evaluation and other incentive structures (Wallace and Rafols, 2016). In this paper, we abstract from this complexity and use publications as the resulting output.

By employing co-word analysis, we identified the main topics of research on rice from a corpus of more than 100,000 publications, distinguishing among (i) plant protection divided into pests and weeds (*plant protection*); (ii) production practices and socio-economic issues ((*production practices*); (iii) plant nutrition and yields ((*yields*); (iv) rice varieties, and traditional genetics (*traditional genetics*); (v) transgenic, molecular biology, and genomics (*genomics*); and (vi) consumption, human nutrition and food technologies (*human consumption*). We then identified variations in research priorities across countries and years by allocating publications to topics.

Second, we compiled a number of proxies for a country's societal demands for research, based on uses and needs related to rice. Because in this case, too, we cannot observe the preferences of single farmers and consumers and how they interact and reach those who set priorities and produce research, we relied on aggregate indicators of demand for rice (from national accounts) and on societal needs related to rice. We refer to these as revealed societal demands.

Third, we estimated the relation between countries' specialisation in different topics over the years, and their revealed demands for rice research and technology.

We found that revealed demands for rice technologies to a limited extent explain a country's revealed research priorities on rice: the alignment is partial. We found a positive correlation between priorities and demands in a number of cases. For instance, countries with lower yields invest more than others in research on yields (and plant nutrition), traditional genetics and production practices.

However, we found a number of misalignments between the focus of rice research and revealed demands, in particular, related to human nutrition. Countries that use rice mainly for food were found to specialise only in production practices and socio-economic issues. Most importantly, these countries do not specialise in human consumption (and food technologies). On the contrary, countries where rice is an important source of caloric intake, specialise significantly less in human consumption and, also, traditional genetics and genomics. It is of concern that genomics, the field of research that has produced the largest share of publication in recent years, is the least related to demands for rice research and technologies. Production practices and socio-economic issues seem to be the research topic most related to social needs. However, in the case of rice technology, it is questionable whether social sciences, alone, can solve some of the main issues related to nutrition demands, climate change problems, changes to biotic and abiotic stresses and increasing population.

There are many reasons for these apparent misalignments, some of which are discussed in Section 2. In future work, we hope to dig deeper into some of these misalignments and study the research prioritisation process and how it is shaped by societal demands and institutional practices. The search for such explanations should also consider the recent shift from public to private research in agriculture (Pardey et al., 2016) (although this applies less to the case of rice). Private research may differ from public research and focus on more profitable topics such as crop genetics, machinery and food processing.

Our results show that a systemic approach to portfolio analysis of agriculture research could help identify research topics and promote research more closely aligned to socio-economic needs. Agriculture and food production are too relevant and too closely related to the satisfaction of basic needs (of consumers and producers) to be confined to a productivity-export led strategy – which our data suggest to be closer to priorities in industrialised and some emerging countries.

This paper has some limitations related to the empirical strategy that we hope to address in future work. First, although CAB's publications coverage is possibly the largest on the subject of rice, publications represent a subset of research outputs. In agricultural research, especially, many research outputs are not accounted for by publications, such as experiments in the field, and a substantial proportion of the research conducted by private companies and public organisations. Second, our account of the research and demand priorities is aggregated at the country level, and refers to observed (revealed) supply and demands, not to actual research priorities and demands. We need a better account and mapping of unpublished knowledge and societal needs.

References

Arndt, Thomas M., Dana G. Dalrymple, and Vernon W. Ruttan, eds, *Resource Allocation and Productivity in National and International Agricultural Research*, Minneapolis: University of Minnesota Press, 1977.

Arthur, W Brian, "Competing technologies, increasing returns and lock-in by historical events," *Economic Journal*, 1989, *99*, 116–131.

_ , "Inductive reasoning and bounded rationality," *American Economic Review*, 1994, *84*, 406.

Arvanitis, Rigas and Yvon Chatelin, "National Scientific Strategies in Tropical Soil Sciences," *Social Studies of Science*, 1988, *18* (1), 113–146.

Balassa, Bela, "Trade Liberalisation and "Revealed" Comparative Advantage1," *The Manchester School*, 1965, *33* (2), 99–123.

Bantilan, M. C. S. and J. D. H. Keatinge, *Considerations for Determining Research Priorities: Learning Cycles and Impact Pathways*, Dordrecht: Springer Netherlands,

Barke, Richard P. and Hank C. Jenkins-Smith, "Politics and Scientific Expertise: Scientists, Risk Perception, and Nuclear Waste Policy," *Risk Analysis*, aug 1993, *13* (4), 425–439.

Binswanger, Hans P., *Measuring the impact of economic factors on the direction of technical change*, Minneapolis: University of Minnesota Press,

Bozeman, Barry and Daniel Sarewitz, "Public values and public failure in US science policy," *Science and Public Policy*, apr 2005, *32* (2), 119–136.

_ **and** _ , "Public Value Mapping and Science Policy Evaluation," *Minerva*, Mar 2011, *49* (1), 1–23.

Brooks, Sally, "Is international agricultural research a global public good? The case of rice biofortification," *Journal of Peasant Studies*, 2011, *38* (1), 67–80.

_ , "Living with Materiality or Confronting Asian Diversity? The Case of Iron-Biofortified Rice Research in the Philippines," *East Asian Science, Technology and Society*, 2011, *5* (2), 173–188.

Bush, Vannevar, "Science: The Endless Frontier," *Transactions of the Kansas Academy of Science*, 1945, *48* (3), 231–264.

Callon, Michel, Jean-Pierre Courtial, William A. Turner, and Serge Bauin, "From translations to problematic networks: An introduction to co-word analysis," *Information (International Social Science Council)*, 1983, *22* (2), 191–235.

Carlisle, Liz and Albie Miles, "Closing the Knowledge Gap: How the USDA Could Tap the Potential of Biologically Diversified Farming Systems," *Journal of Agriculture, Food Systems, and Community Development*, 2016, *3* (4), 219–225.

Cassi, Lorenzo, Agénor Lahatte, Ismael Rafols, Pierre Sautier, and Élisabeth de Turckheim, "Improving fitness: Mapping research priorities against societal needs on obesity," *Journal of Informetrics*, 2017, *11* (4), 1095 – 1113.

Chataway, Joanna, Joyce Tait, and David Wield, "Understanding company R&D strategies in agro-biotechnology: trajectories and blind spots," *Research Policy*, 2004, *33* (6), 1041 – 1057.

Cowan, Robin and Philip Gunby, "Sprayed to Death: Path Dependence, Lock-in and Pest Control Strategies," *The Economic Journal*, may 1996, *106* (436), 521–542.

Dalrymple, Dana G., "Evaluating the Impact of Internaitonal Research on Wheat and Rice Production in the Developing Countries," in Vernon W. Ruttan Thomas M. Arndt, Dana G. Dalrymple, ed., *Resource Allocation and Productivity in National and International Agricultural Research*, Minneapolis: University of Minnesota Press, 1977.

_ , "Setting the agenda for science and technology in the public sector: the case of international agricultural research," *Science and Public Policy*, 2006, *33* (4), 277–290.

David, Paul A., "The mechanization of reaping in the ante-bellum Midwest," in Robert William Fogel and Stanley L. Engerman, eds., *The Reinterpretation of American Econmic History*, Harper & Row, 1971, chapter 16, pp. 214–227.

_ **and Warren C Sanderson**, "Making Use of Treacherous Advice: Cognitive Progress, Bayesian Adaptation, and the Tenacity of Unreliable Knowledge," in John N Drobak and John V C Nye, eds., *The Frontiers of the New Institutional Economics*, San Diego: Academic Press, 1997, chapter 14, pp. 305–366.

de Janvry, Alain, *Inducement of and Institutional Innovations: An Interpretative Framework*, Minneapolis: University of Minnesota Press,

_ , "Social structure and biased technical change in Argentine agriculture," in Hans Binswanger and Vernon Ruttan, eds., *Induced Innovation: Technology, Institutions and Development*, Baltimore: Johns Hopkins University Press, 1978.

Dosi, Giovanni, "Technological paradigms and technological trajectories : A suggested interpretation of the determinants and directions of technical change," *Research Policy*, 1982, *11(3)*, 147–162.

_ **and Richard R. Nelson**, "The Evolution of Technologies: An Assessment of the State-of-the-Art," *Eurasian Business Review*, 2013, *3* (1), 3–46.

Driscoll, John C. and Aart C. Kraay, “Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data,” *The Review of Economics and Statistics*, November 1998, *80* (4), 549–560.

Eisenstein, Michael, “Biotechnology: Against the grain,” *Nature*, oct 2014, *514* (7524), S55–S57.

Ely, Adrian, Patrick Van Zwanenberg, and Andrew Stirling, “Broadening out and opening up technology assessment: Approaches to enhance international development, co-ordination and democratisation,” *Research Policy*, 2013, (0), –.

Evans, James A., Jae-Mahn Shim, and John P. A. Ioannidis, “Attention to Local Health Burden and the Global Disparity of Health Research,” *PLoS ONE*, 04 2014, *9* (4), e90147.

Evenson, R. E. and D. Gollin, “Assessing the Impact of the Green Revolution, 1960 to 2000,” *Science*, 2003, *300* (5620), 758–762.

Fischer, Klara and Flora Hajdu, “Does raising maize yields lead to poverty reduction? A case study of the Massive Food Production Programme in South Africa,” *Land Use Policy*, 2015, *46* (0), 304 – 313.

Fontana, Roberto, Alessandro Nuvolari, and Bart Verspagen, “Mapping technological trajectories as patent citation networks. An application to data communication standards,” *Economics of Innovation and New Technology*, 2009, *18* (4), 311–336.

Freeman, Christopher, “Innovation, Changes of Techno-Economic Paradigm and Biological Analogies in Economics,” *Revue économique*, mar 1991, *42* (2), 211.

_ , “The “National System of Innovation” in Historical Perspective,” *Cambridge Journal of Economics*, 1995, *19.*

Frickel, S., S. Gibbon, J. Howard, J. Kempner, G. Ottinger, and D. J. Hess, “Undone Science: Charting Social Movement and Civil Society Challenges to Research Agenda Setting,” *Science, Technology & Human Values*, oct 2009, *35* (4), 444–473.

Fuchs, Doris and Katharina Glaab, “Material power and normative conflict in global and local agrifood governance: The lessons of ‘Golden Rice’ in India,” *Food Policy*, 2011, *36* (6), 729 – 735. ¡ce:title¿Between the Global and the Local, the Material and the Normative: Power struggles in India’s Agrifood System¡/ce:title¿.

Fuller, Dorian Q, Robin G. Allaby, and Chris Stevens, “Domestication as innovation: the entanglement of techniques, technology and chance in the domestication of cereal crops,” *World Archaeology*, 2010, *42* (1), 13–28.

Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, and M. Trow, *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, SAGE Publications, 1994.

Gillum, Leslie A., Christopher Gouveia, E. Ray Dorsey, Mark Pletcher, Colin D. Mathers, Charles E. McCulloch, and S. Claiborne Johnston, "NIH Disease Funding Levels and Burden of Disease," *PLOS ONE*, 02 2011, *6* (2), 1–9.

Gläser, Jochen and Grit Laudel, "Governing Science: How Science Policy Shapes Research Content," *European Journal of Sociology*, 004 2016, *57* (1), 117–168.

Graff, G, "Observing Technological Trajectories in Patent Data:Empiricalmethods To Study the Emergence and Growth of New Technologies," *American Journal of Agricultural Economics*, dec 2003, *85* (5), 1266–1274.

Hannerz, F. and A. Lotsch, "Assessment of remotely sensed and statistical inventories of African agricultural fields," *International Journal of Remote Sensing*, 2008, *29* (13), 3787–3804.

Herdt, R.W., "People, institutions, and technology: A personal view of the role of foundations in international agricultural research and development 1960-2010," *Food Policy*, 2012, *37* (2), 179–190. cited By 8.

Herring, Ronald J., "The genomics revolution and development studies: Science, poverty and politics," *Journal of Development Studies*, 2007, *43* (1), 1–30.

Hess, David J., Sulfikar Amir, Scott Frickel, Daniel Lee Kleinman, Kelly Moore, and Logan D. A. Williams, "Structural Inequality and the Politics of Science and Technology," in Clark A. Mille Ulrike Fel and, Rayvon Fouché and and Laurel Smith-Doerr, eds., *The handbook of science and technology studies*, fourth ed., Cambridge, MA: MIT Press, 2017, chapter 11, pp. 319–347.

Hogg, D., *Technological Change In Agriculture: Locking in to Genetic Uniformity*, Palgrave Macmillan UK, 2000.

Hsieh, S.C. and V.W. Ruttan, "Environmental, Technological, and Institutional Factors in the Growth of Rice Production: Philippines, Thailand, and Taiwan," in "Can Economic Growth Be Sustained?: The Collected Papers of Vernon W. Ruttan and Yujiro Hayami," Oxford University Press, 2015.

IFPRI, "2012 Global Food Policy Report," Technical Report, International Food Policy Research Institute (IFPRI), Washington, DC 2013.

Jenkins, Mark and Steven Floyd, "Trajectories in the evolution of technology: A multi-level study of competition in Formula 1 racing," *Organization Studies*, 2001, *22* (6), 945–969.

Johnstone, Phil and Andy Stirling, "Comparing Nuclear Power Trajectories in Germany And The UK: From Regimes' to Democracies' in Sociotechnical Transitions and Discontinuities," 2015.

Joly, Pierre-Benot, Ariane Gaunand, Laurence Colinet, Philippe Lardo, Stphane Lemari, and Mireille Matt, "ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization," *Research Evaluation*, 2015, *24* (4), 440.

Kahlon, A.S, P.N Saxena, H.K. Bal, and D. Jha, *Returns to investment in agricultural research in India*, Minneapolis: University of Minnesota Press,

Kelley, T.G., J.G. Ryan, and B.K. Patel, "Applied participatory priority setting in international agricultural research: Making trade-offs transparent and explicit," *Agricultural Systems*, 1995, *49* (2), 177 – 216.

Kinge, Jonas Minet, Ingrid Roxrud, Stein Emil Vollset, Vegard Skirbekk, and John-Arne Røttingen, "Are the Norwegian health research investments in line with the disease burden?," *Health Research Policy and Systems*, Nov 2014, *12* (1), 64.

Klerkx, Laurens and Cees Leeuwis, "Institutionalizing end-user demand steering in agricultural R&D: Farmer levy funding of R&D in The Netherlands," *Research Policy*, apr 2008, *37* (3), 460–472.

_ **and** _ , "Matching demand and supply in the agricultural knowledge infrastructure: Experiences with innovation intermediaries," *Food Policy*, 2008, *33* (3), 260 – 276.

_ **, Barbara van Mierlo, and Cees Leeuwis**, *Evolution of systems approaches to agricultural innovation: concepts, analysis and interventions*, Dordrecht: Springer Netherlands,

Kline, S.J. and N. Rosenberg, "An overview of innovation," in R. Landau & N. Rosenberg, ed., *The Positive Sum Strategy: Harnessing Technology for Economic Growth*, Washington, D.C.: National Academy Press, 1986, p. 275305.

Kolady, Deepthi Elizabeth, David J. Spielman, and Anthony Cavalieri, "The Impact of Seed Policy Reforms and Intellectual Property Rights on Crop Productivity in India," *Journal of Agricultural Economics*, 2012, *forthcoming.*

Leydesdorff, Loet and Iina Hellsten, "Measuring the meaning of words in contexts: An automated analysis of controversies about 'Monarch butterflies,' 'Frankenfoods,' and 'stem cells'," *Scientometrics*, 2006, *67* (2), 231–258.

Ligon, Ethan and Elisabeth Sadoulet, "Estimating the effects of aggregate agricultural growth on the distribution of expenditures," Department of Agricultural & Resource Economics, UC Berkeley, Working Paper Series, Department of Agricultural & Resource Economics, UC Berkeley 2011.

MacKenzie, Donald, "Economic and Sociological explanations of Technological Change," in "Knowing machines: essays on technical change," {MIT} Press, jul 1998, pp. 49–65.

Marin, Anabel, Lilia Stubrin, and Patrick van Zwanenberg, "Developing capabilities in the seed industry: which direction to follow?," SPRU Working Paper Series 2014-12, SPRU - Science and Technology Policy Research, University of Sussex June 2014.

Matt, M., A. Gaunand, P-B. Joly, and L. Colinet, "Opening the black box of impact – Ideal-type impact pathways in a public agricultural research organization," *Research Policy*, 2017, *46* (1), 207 – 218.

McGuire, Shawn J., "Path-dependency in plant breeding: Challenges facing participatory reforms in the Ethiopian Sorghum Improvement Program," *Agricultural Systems*, 2008, *96* (1), 139 – 149.

McNie, Elizabeth C., "Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature," *Environmental Science & Policy*, feb 2007, *10* (1), 17–38.

Miller, Thaddeus R. and Mark W. Neff, "De-Facto Science Policy in the Making: How Scientists Shape Science Policy and Why it Matters (or, Why STS and STP Scholars Should Socialize)," *Minerva*, jul 2013, *51* (3), 295–315.

Mina, A., R. Ramlogan, G. Tampubolon, and J.S. Metcalfe, "Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge," *Research Policy*, jun 2007, *36* (5), 789–806.

Mokyr, Joel, "Evolutionary phenomena in technological change," in John Michael Ziman, ed., *Technological Innovation as an Evolutionary Process*, Cambridge: Cambridge University Press, 2000, chapter 5, pp. 52–65.

Morooka, K., M.M. Ramos, and F.N. Nathaniel, "A bibliometric approach to interdisciplinarity in Japanese rice research and technology development," *Scientometrics*, 2014, *98* (1), 73–98. cited By 1.

Mowery, David and Nathan Rosenberg, "The influence of market demand upon innovation: a critical review of some recent empirical studies," *Research Policy*, 1979, *8* (2), 102 – 153.

Nelson, Richard R., "On the uneven evolution of human know-how," *Research Policy*, jun 2003, *32* (6), 909–922.

_ , "The Moon and the Ghetto revisited," *Science and Public Policy*, nov 2011, *38* (9), 681–690.

Norton, George W., Philip G. Pardey, and Julian M. Alston, "Economic Issues in Agricultural Research Priority Setting," *American Journal of Agricultural Economics*, 1992, *74* (5), 1089–1094.

Orr, Alastair, "Why were So Many Social Scientists Wrong about the Green Revolution? Learning from Bangladesh," *Journal of Development Studies*, 2012, *48* (11), 1565–1586.

Parayil, Govindan, "Mapping technological trajectories of the Green Revolution and the Gene Revolution from modernization to globalization," *Research Policy*, June 2003, *32* (6), 971–990.

Pardey, Philip G., Connie Chan-Kang, Steven P. Dehmer, and Jason M. Beddow, "Agricultural R&D is on the move," *Nature*, 2016, *537*, 301–303.

Pavitt, Keith, "The social shaping of the national science base," *Research Policy*, 1998, *27* (8), 793 – 805.

Pinch, T. J. and W. E. Bijker, "The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other," *Social Studies of Science*, aug 1984, *14* (3), 399–441.

Pingali, Prabhu L., Scott D. Rozelle, and Roberta V. Gerpacio, "The Farmer's Voice in Priority Setting: A CrossCountry Experiment in Eliciting Technological Preferences," *Economic Development and Cultural Change*, 2001, *49* (3), 591–609.

Pinstrup-Andersen, Per and David Franklin, "A Systems Approach to Agricultural Research Resource Allocation in Developing CountrIes," in Vernon W. Ruttan Thomas M. Arndt, Dana G. Dalrymple, ed., *Resource Allocation and Productivity in National and International Agricultural Research*, Minneapolis: University of Minnesota Press, 1977.

Plutzer, Eric, Ardith Maney, and R. O'Connor, "Ideology and elites' perceptions of the safety of new technologies.," *American Journal of Political Science*, 1998, *42* (1), 190–209.

Polanyi, Michael, "The Republic of Science: Its Political and Economic Theory," *Minerva*, mar 2000, *38* (1), 1–21.

Possas, Mario Luiz, Sergio Salles-Filho, and JosMaria da Silveira, "An evolutionary approach to technological innovation in agriculture: some preliminary remarks," *Research Policy*, 1996, *25* (6), 933 – 945.

Rafols, Ismael and Alfredo Yegros, "Is Research Responding to Health Needs?," *Social Observatory of La Caixa Foundation*, March 2018, (March).

_ , Tommaso Ciarli, and Diego Chavarro, "Under-Reporting Research Relevant to Local Needs in the Global South. Database Biases in the Representation of Knowledge on Rice," in "International Conference on Scientometrics and Informetrics" SPRU, University of Sussex International Conference on Scientometrics and Informetrics Istanbul 2015.

Raitzer, D.A. and G.W Norton, *Introduction to Prioritizing Agricultural Research for Development*, Vol. Prioritizing Agricultural Research for Development: Experiences and Lessons of *Cabi Series*, CABI,

Ray, Deepak K., Nathaniel D. Mueller, Paul C. West, and Jonathan A. Foley, "Yield Trends Are Insufficient to Double Global Crop Production by 2050," *PLOS ONE*, 06 2013, *8* (6), 1–8.

Roberts, Royston M., *Serendipity: Accidental Discoveries in Science*, Wiley, June 1989.

Römheld, Volker and Ernest A. Kirkby, "Research on potassium in agriculture: needs and prospects," *Plant and Soil*, Oct 2010, *335* (1), 155–180.

Rotolo, Daniele, Ismael Rafols, Michael M. Hopkins, and Loet Leydesdorff, "Strategic intelligence on emerging technologies: Scientometric overlay mapping," *Journal of the Association for Information Science and Technology*, January 2017, *68* (1), 214–233.

_ , Ismale Rafols, Michael M. Hopkins, and Loet Leydesdorff, "Building a Multi-Perspective Scientometric Approach on Tentative Governance in Emerging Technologies," Conference proceeding, ISSI Conference, Vienna 2013.

Ruttan, Vernon W, "Induced Innovation, Evolutionary Theory and Path Dependence: Sources of Technical Change," *Economic Journal*, 1997, *107* (444), 1520–1529.

Salton, Gerard, Edward A. Fox, and Harry Wu, "Extended Boolean Information Retrieval," *Communications of the ACM*, November 1983, *26* (11), 1022–1036.

Samuelson, P. A., "A Note on the Pure Theory of Consumer's Behaviour," *Economica*, 1938, *5* (17), 61–71.

Sarewitz, Daniel, *Frontiers Of Illusion: Science, Technology, and the Politics of Progress*, Temple University Press, 2010.

_ and Roger a. Pielke, "The neglected heart of science policy: reconciling supply of and demand for science," *Environmental Science & Policy*, feb 2007, *10* (1), 5–16.

Schmookler, J., *Invention and economic growth*, Harvard University Press, 1966.

Schuhmacher, Alexander, Oliver Gassmann, and Markus Hinder, "Changing R&D models in research-based pharmaceutical companies," *Journal of Translational Medicine*, apr 2016, *14* (1).

Shiva, Vandana and Tom Crompton, "Monopoly and Monoculture: Trends in Indian Seed Industry," *Economic and Political Weekly*, 1998, *33* (39), pp. A137–A141+A144–A151.

Smith, Adrian, Andy Stirling, and Frans Berkhout, "The governance of sustainable socio-technical transitions," *Research Policy*, dec 2005, *34* (10), 1491–1510.

Spaapen, Jack and Leonie van Drooge, "Introducing productive interactions in social impact assessment," *Research Evaluation*, 2011, *20* (3), 211–218.

Sumberg, James, "The logic of fodder legumes in Africa," *Food Policy*, 2002, *27* (3), 285 – 300.

_ , John Thompson, and Philip Woodhouse, "Why agronomy in the developing world has become contentious," *Agriculture and Human Values*, 2013, *30* (1), 71–83.

Thompson, John and Ian Scoones, "Addressing the dynamics of agri-food systems: an emerging agenda for social science research," *Environmental Science & Policy*, 2009, *12* (4), 386 – 397.

Touzard, Jean-Marc and Ludovic Temple, "Sécurisation alimentaire et innovations dans lagriculture et lagroalimentaire: vers un nouvel agenda de recherche? Une revue de la littérature," *Cahiers Agricultures*, 2012, *21* (5), 293–301.

van Eck, Nees Jan and Ludo Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, 2010, *84* (2), 523–538.

_ , _ , Ed C. M. Noyons, and Reindert K. Buter, "Automatic term identification for bibliometric mapping," *Scientometrics*, 2010, *82* (3), 581–596.

_ , _ , Rommert Dekker, and Jan van den Berg, "A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS," *Journal of the American Society for Information Science and Technology*, 2010, *61* (12), 2405–2416.

Vanloqueren, Gaëtan and Philippe V. Baret, "Why are ecological, low-input, multi-resistant wheat cultivars slow to develop commercially? A Belgian agricultural lock-in case study," *Ecological Economics*, 2008, *66* (2), 436 – 446.

_ and _ , "How agricultural research systems shape a technological regime that develops genetic engineering but locks out agroecological innovations," *Research Policy*, 2009, *38* (6), 971 – 983.

Verspagen, Bart, "Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research," *Advances in Complex Systems*, 2007, *10* (01), 93–115.

Wallace, Matthew L. and Ismael Rafols, "Research Portfolio Analysis in Science Policy: Moving from Financial Returns to Societal Benefits," *Minerva*, Apr 2015, *53* (2), 89–115.

_ **and** _ , "Shaping the Agenda of a Grand Challenge: Institutional Mediation of Priorities in Avian Influenza Research," 2016.

Waltman, Ludo, Nees Jan van Eck, and Ed C.M. Noyons, "A unified approach to mapping and clustering of bibliometric networks," *Journal of Informetrics*, 2010, *4* (4), 629 – 635.

Welch, Ross M and Robin D Graham, "A new paradigm for world agriculture: meeting human needs: Productive, sustainable, nutritious," *Field Crops Research*, 1999, *60* (1), 1 – 10.

Wolff, H. and G. Recke, "Path dependence and implementation strategies for integrated pest management.," *Quarterly Journal of International Agriculture*, 2000, *39* (2), 149–171.

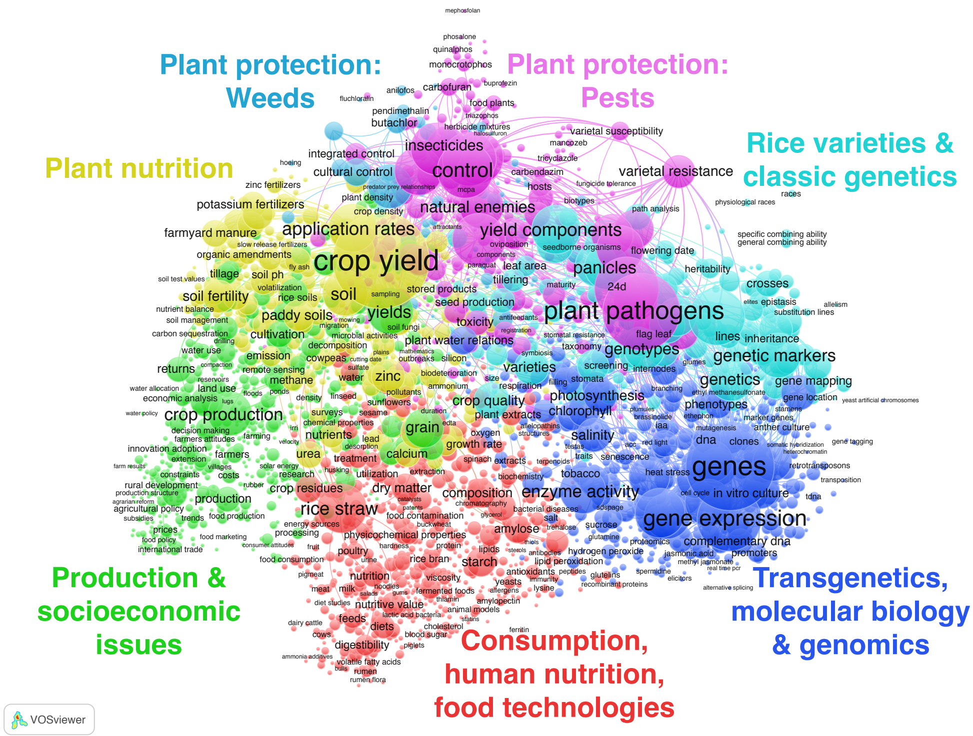
Woolston, Chris, "Rice," *Nature*, 10 2014, *514* (7524), S49–S49.

Wright, Brian D., "Grand missions of agricultural innovation," *Research Policy*, 2012, *41* (10), 1716 – 1728. The need for a new generation of policy instruments to respond to the Grand Challenges.

Yaqub, Ohid, "Serendipity: Towards a taxonomy and a theory," *Research Policy*, 2018, *47* (1), 169 – 179.

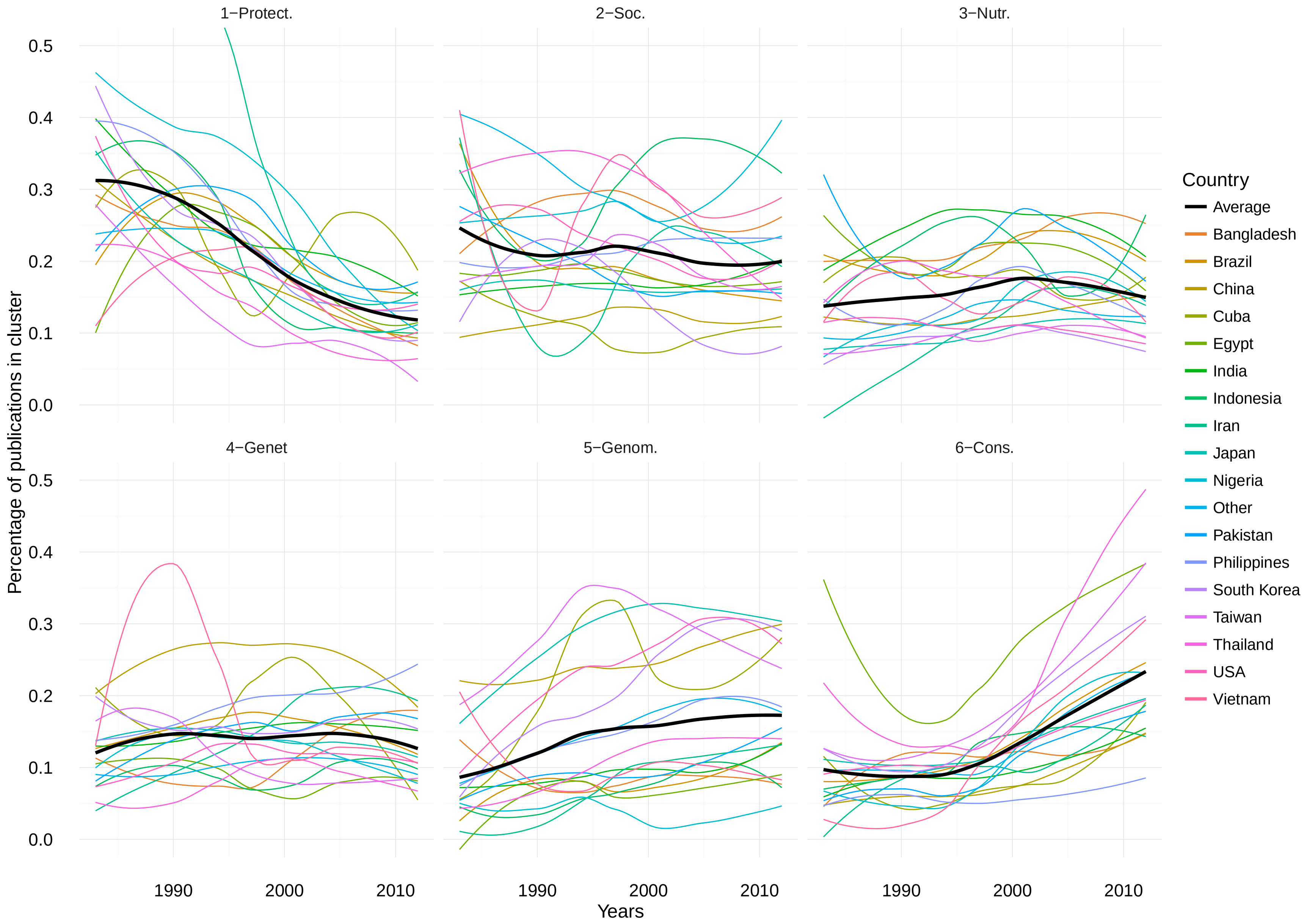
A Figures

A.1 Paper



Notes: each colour represent a cluster of terms (descriptors assigned to publications in CAB), which identifies a research topic. The clustering was computed using the VOSviewer algorithm based on the co-occurrence matrix computed as described in Section 4. Each node represents a term. To improve visualisation, terms are reported only for the main nodes. The size of a term (node) is proportional to the relative number of publications that contain that term. The presence of an edge indicates that two terms are similar (they appear frequently in the same publications). To improve visualisation, we report only the top 500 edges. Geographic distance represents the proximity between terms: more distant clusters suggest that terms in these clusters do not appear together in the same publications frequently.

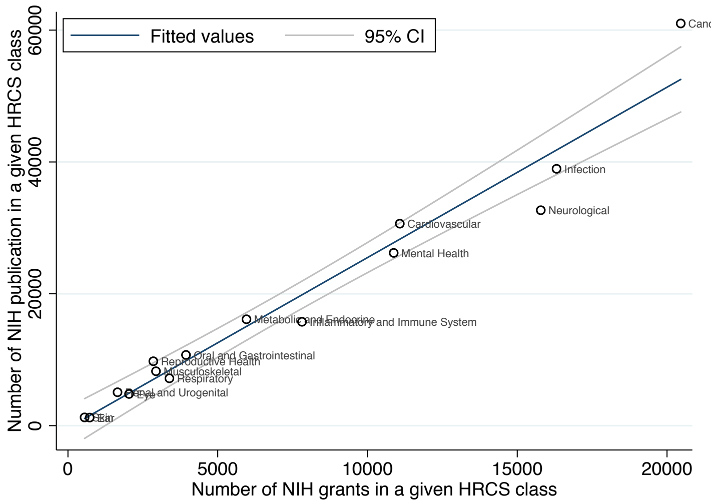
Figure 1: Global map of rice research (1983-2012). Co-word clustering of publications from the CAB database.



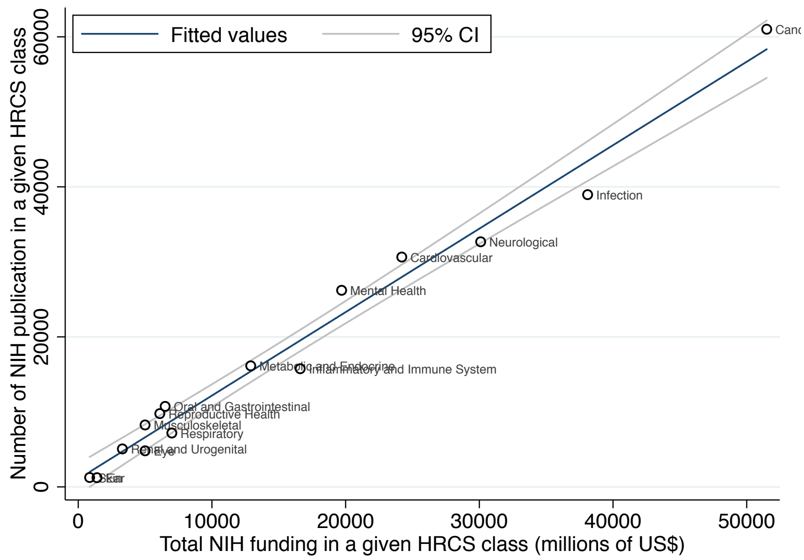
Notes: each series represents the locally smoothed regression (LOESS) of the percentage of papers published in a topic. The topics are derived from the clustering exercise plotted in the global map (Figure 1). These are: *1-Protect*: plant protection (weeds and pests); *2-Soc*: production practices and socio economic issues; *3-Nutr*: plant nutrition and yields; *4-Genet*: rice varieties, and classical genetics; *5-Genom*: transgenics, molecular biology, and genomics; *6-Cons*: consumption, human nutrition and food. The percentage of papers is computed using fractional counting.

Figure 2: Differences in technological trajectory across countries.

A.2 Web Appendix



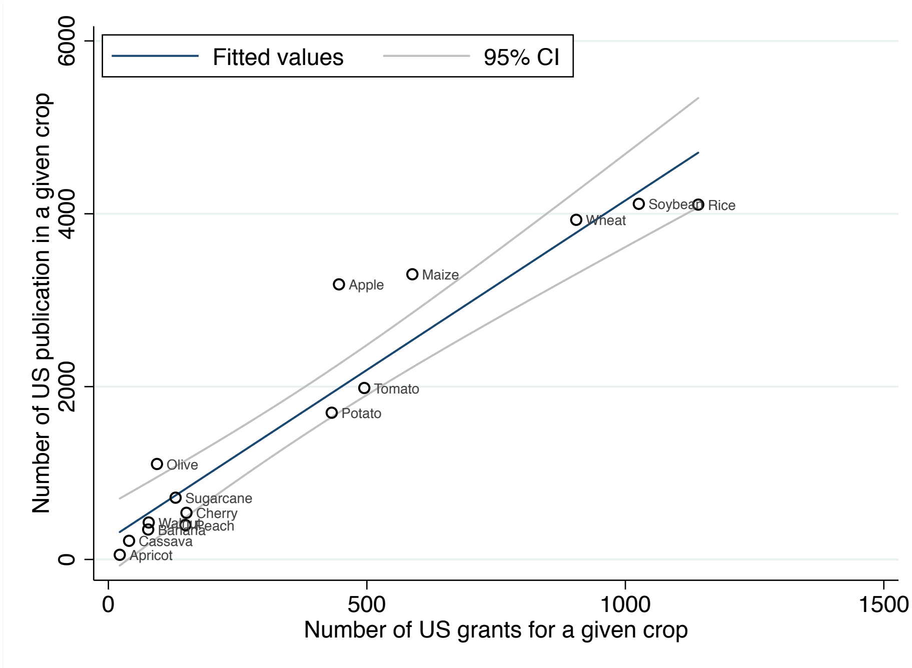
(a) Number of NIH grants



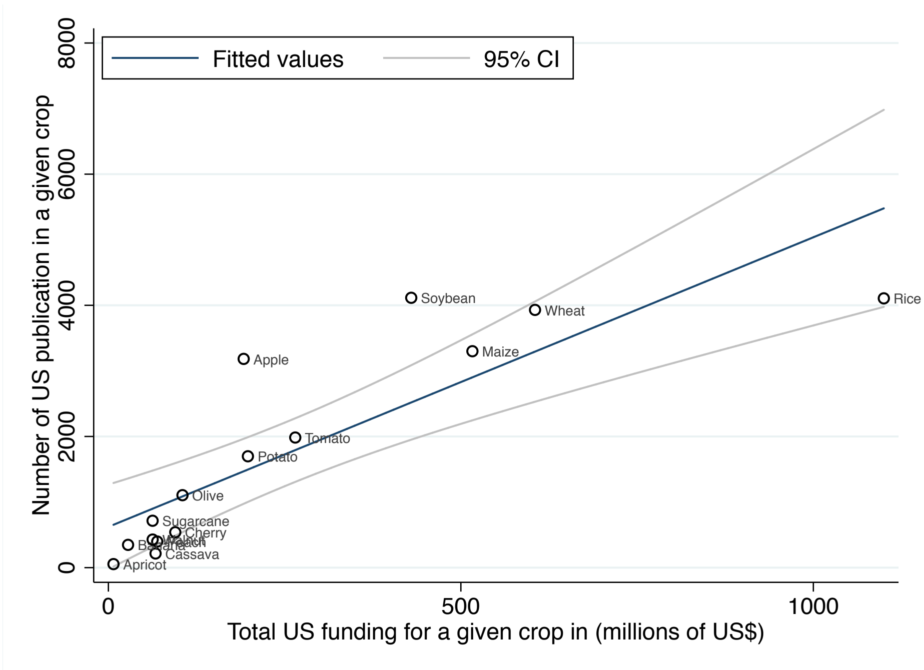
(b) Total NIH funding

Notes: for each illness, we retrieved data on funding from the National Institute of Health (NIH) and the publications that acknowledge NIH grants between 2014-17. We estimated the linear fit between the total number of grants per illness and the total number of publications on that illness, in the same period (panel 3a); and the relation between the total amount of funds (in US$ m) in grants for research on an illness, and the total number of publication on that illness in the same period (panel 3b). Data were collated by Digital Science and made available via the online tool Dimensions. Information on the funder for each publication was retrieved from the publications' acknowledgement. The straight blue line is the best linear fit between funding and publications and the grey lines represent its confidence intervals.

Figure 3: Relation between NIH funding (inputs) and publications (outputs) across disease classes (HRCS) (2014-17)



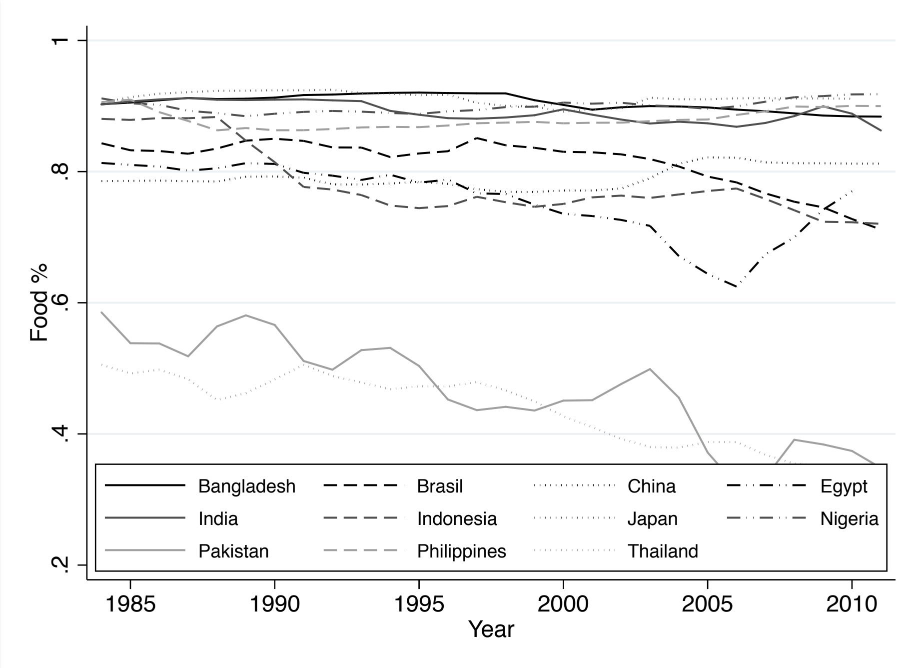
(a) Number of grants



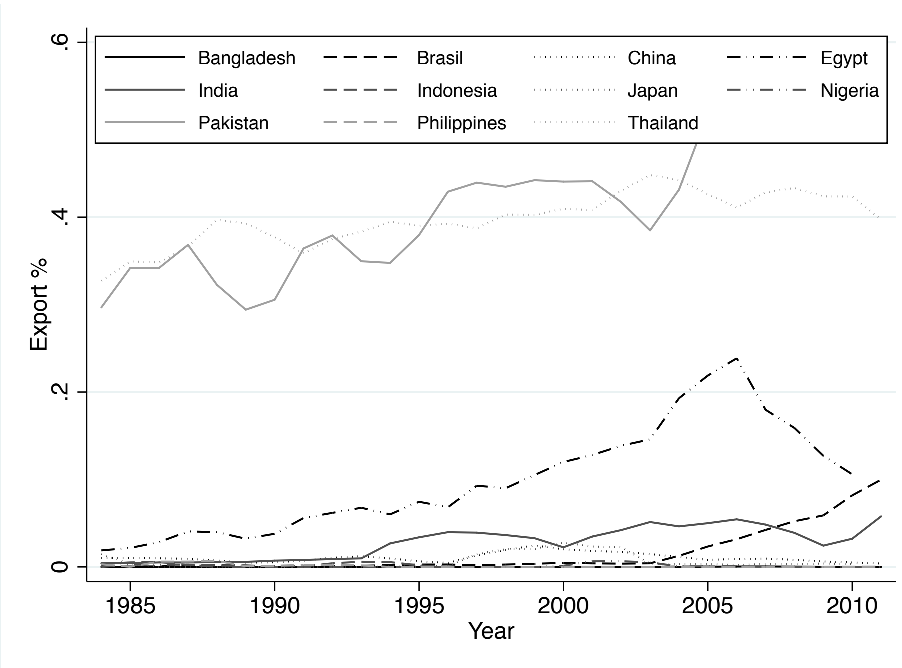
(b) Total funding

Notes: for each crop, we retrieved data on funding in the US, and the publications that acknowledge US grants between 2013-17. We estimated the linear fit between the total number of grants per crop and the total number of publications on that crop in the same period (panel 4a); and the relation between the total amount of funds (in US$m) in grants for research on a crop, and the total number of publication on that crop in the same period (panel 4b). Data were collated by Digital Science and made available via the online tool Dimensions. Information on the funder for each publication is retrieved from the acknowledgement. The straight blue line is the best linear fit between funding and publications, and the grey lines represent its confidence intervals.

Figure 4: Relation between US funding (inputs) and publications (outputs) across crops (2013-17)



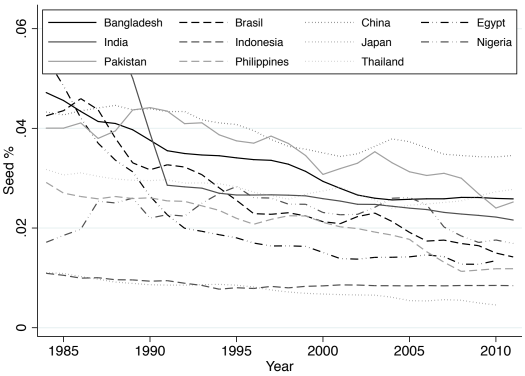
(a) Food consumption



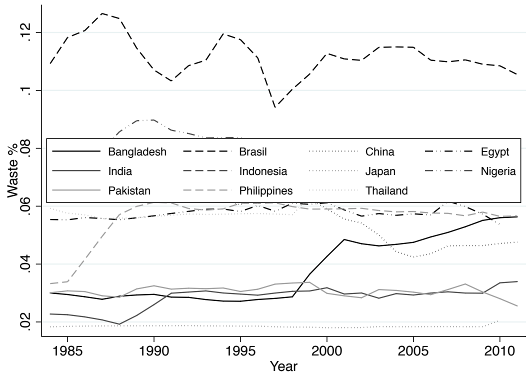
(b) Export

Notes: Each series represents the time change in the percentages of rice used for food (5a) and exported (5b) in a given country. Data from national accounts collected by FAOSTAT.

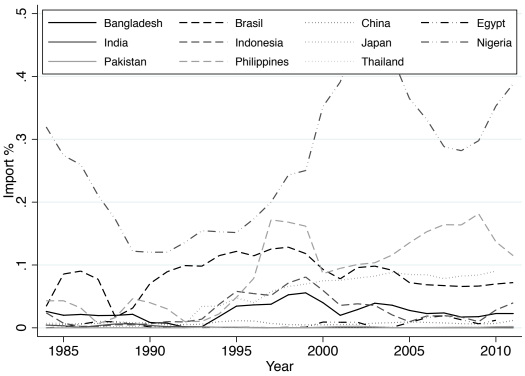
Figure 5: Changes in the relative use of rice produced in a given country: food consumption (5a) and export (5b)



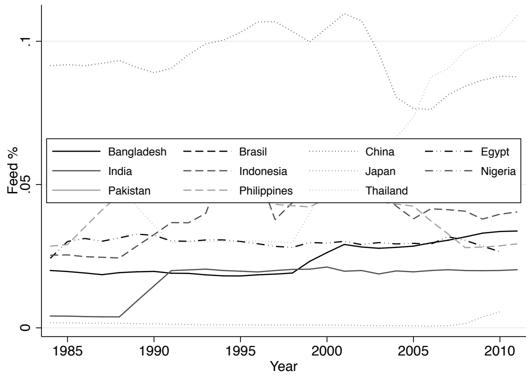
(a) Seeds



(b) Wasted



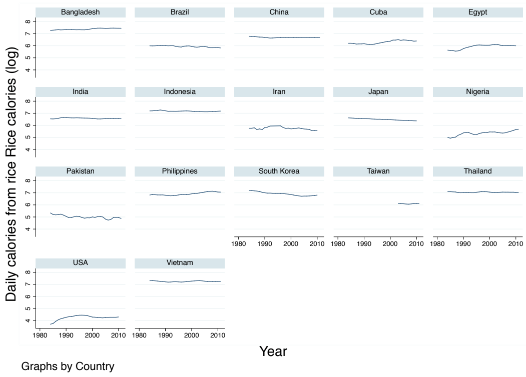
(c) Imported



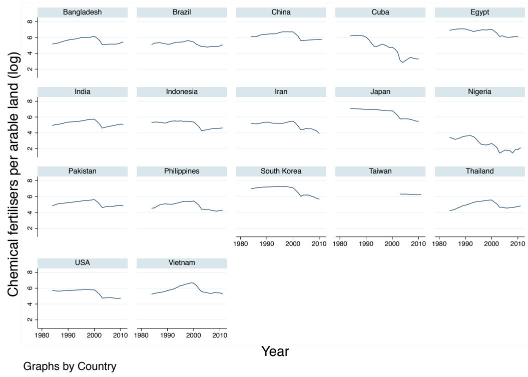
(d) Animal feed

Notes: Each series represents the time change in the percentages of rice used for seeds (6a), wasted (6b), imported (6c) and animal feed (6d) in a given country. Data from national accounts collected by FAOSTAT.

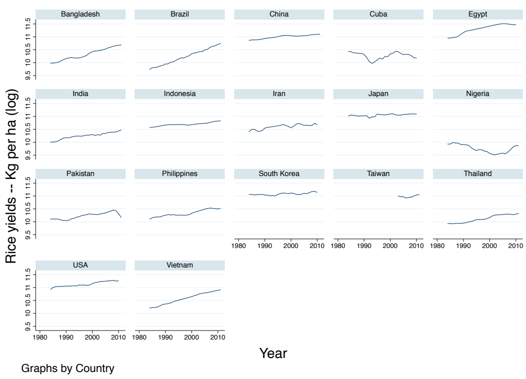
Figure 6: Changes in the relative use of rice produced in a given country: seeds (6a), wasted (6b), imported (6c), and animal feed (6d)



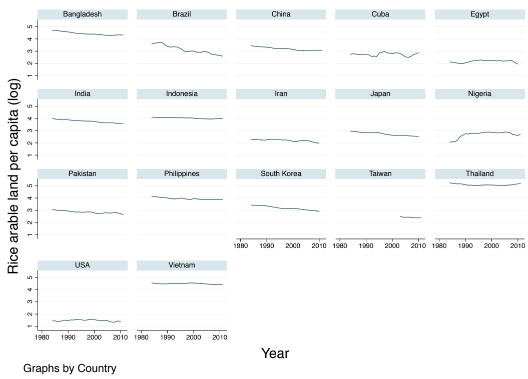
(a) Caloric intake from rice



(b) Chemical fertilisers



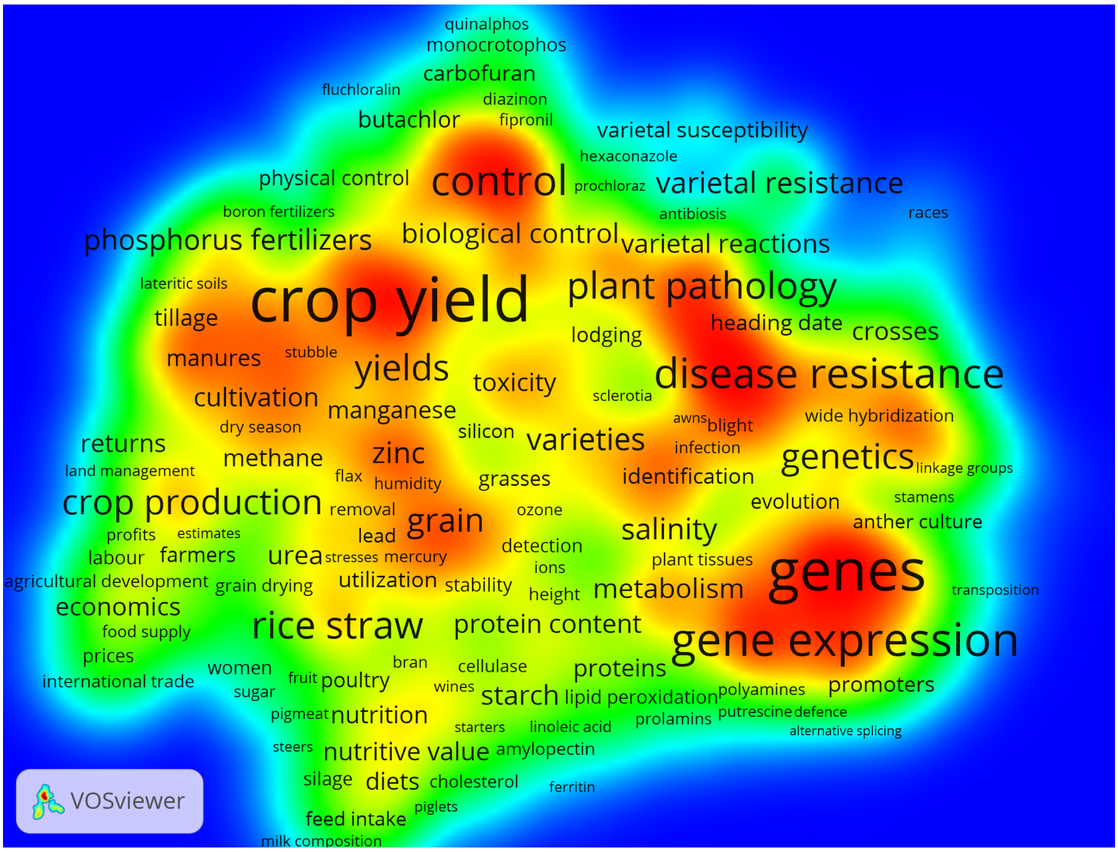
(c) Rice yields



(d) Area cultivated with rice

Notes: Each series represents the time change, for each country, in one of the proxies used to measure revealed demands related to farmers' and consumers' potential needs: the amount of per capita caloric intake from the consumption of rice in logs (7a); the amount of chemical fertilisers used per arable land in logs (7b); rice yields measured as kilograms per hectare, in logs (7c), and the total amount of hectares of arable land devoted to rice cultivation, in logs (7d). Data from national accounts collected by FAOSTAT.

Figure 7: Changes in proxies of revealed demands related to farmers' and consumers' potential needs: caloric intake from rice (7a), fertilisers use (7b), yields (7c), and area dedicated to rice (7d)



Notes: the colour scale represents the relative frequency of papers containing a specific term, from blue (low frequency) to red (high frequency). The clustering was computed using the VOSviewer algorithm based on the co-occurrence matrix computed as described in Section 4. Each node represents a term. To improve visualisation, terms are reported only for the main nodes. The size of the text is proportional to the relative number of publications that contain that term. Geographic distance represents proximity among terms: more distant clusters suggest that terms from these clusters do not appear together in the same publication frequently.

Figure 8: Global density map of rice research (1983-2012). Co-word clustering of publications from the CAB database

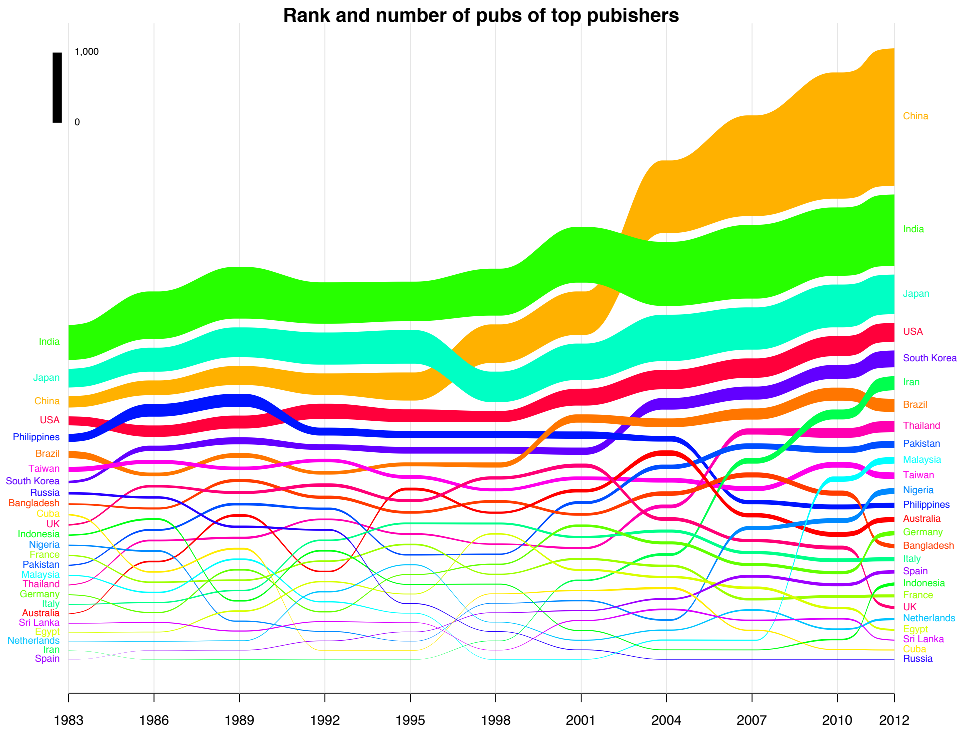


Figure 9: Changes in the rankings and total number of yearly publications for the top country by number of publications/rice production.

B Tables

B.1 Paper

(a)	(b)	(c)	(d)
Variable	**Observed**	**Revealed demand**	**Expected topic**
export	% of rice production exported	No need for consumption	! Human consumption; yields; trad. genetics; genomics
food	% of rice production consumed	Improved consumption	Human consumption; trad. genetics; genomics
seeds	% of rice production stored for future planting	Access to seeds (or seed technology)	Traditional genetics
waste	% of rice production wasted	Storage (related to consumption), and distribution	Production practices; human consumption
imports	% of rice production imported	All related to increased production	Yields; plant protection; traditional genetics; human consumption
processed	% of rice production used in the food industry	Yields and nutritional features	Human consumption
Ricecalories	per capita human intake of calories from rice	Nutrition	Human consumption
Riceyields	productivity	Improved productivity	Yields
Fertilisers	amount of chemical fertilisers used per crop area	Improved soil nutrition	Yields
Ricearea	per capita amount of area devoted to rice cultivation	Reliance on rice (farmers and consumers)	All
Undernourished	% of undernourished population	Nutrition	Human consumption
Pesticides	amount of chemical pesticides used per crop area	Improved pest control	Plant protection

Notes: column (a) reports the variable used to measure a given revealed demand: the rows 1-7 distinguish among different uses of rice in a country (together with *other* the sum of the different percentage is 1), and the remaining rows are further indicators of revealed demands related to needs; column (b) describes the variable in a few words; column (c) describes the societal demands proxied by the variable; column (d) lists the topics that we expect to be most related to societal demand: in some cases more than one, in other cases, in the same row, we indicate with ! topic(s) that we do not expect to be related to societal demand – some of the research topics are not listed here because they could be relevant to multiple societal demands and not one or more specific demands, e.g., genomics. Topics are computed from the publication corpus and are described in Section 5.1.

Table 1: Revealed demands: measurement and relation with research topics

VARIABLES	(1) Plant protect.	(2) Yields	(3) Trad. genetics	(4) Genomics	(5) Human cons.	(6) Prod. practices
Rev. demands						
Export %	0.07***	-0.01	-0.01	-0.09	0.01	-0.06***
	(0.02)	(0.02)	(0.03)	(0.06)	(0.03)	(0.02)
Food %	0.05	0.45	-0.38	-1.40*	-0.06	1.08**
	(0.35)	(0.28)	(0.39)	(0.69)	(0.45)	(0.40)
Seed %	-0.22	0.05	0.46**	-0.55	0.28**	-0.24***
	(0.15)	(0.18)	(0.17)	(0.39)	(0.12)	(0.08)
Waste%	-0.24	0.12	0.72***	-0.19	-0.53***	-0.15
	(0.14)	(0.13)	(0.07)	(0.30)	(0.15)	(0.10)
Import %	-0.04	0.02	-0.03	0.02	-0.00	-0.04*
	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)
Controls						
Education	-0.18**	-0.29***	-0.01	0.17	0.05	0.17**
	(0.08)	(0.09)	(0.06)	(0.19)	(0.18)	(0.06)
Agr. VA	-0.20	0.29*	-0.24	0.07	0.37*	-0.07
	(0.14)	(0.15)	(0.19)	(0.36)	(0.20)	(0.10)
Agr. VA %	0.15	-0.11	-0.47**	-0.11	0.69***	0.14
	(0.20)	(0.17)	(0.19)	(0.31)	(0.15)	(0.17)
Cereals Y.	0.47*	0.17	0.64**	-0.58	0.05	-0.54***
	(0.26)	(0.17)	(0.23)	(0.48)	(0.23)	(0.19)
Constant	-3.99	-2.20	1.27	0.13	-5.46**	3.26
	(3.97)	(2.34)	(2.60)	(4.87)	(2.22)	(2.44)
Observations	303	303	303	303	303	303
Number of groups	11	11	11	11	11	11
within R-squared	0.393	0.292	0.364	0.131	0.275	0.473
F	87.91	160.5	879.3	444.9	1964	533.6
Prob > F	0	0	0	0	0	0
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Notes: results from a fixed effects panel data regression. The dependent variables in the different columns measure the relative specialisation index (in one period/country) for publications in one of the six topics identified in Figure 1 and discussed in Section 5.1. All measures of relative specialisation are computed as three year averages. Revealed demand variables measure the relative use of total rice production: that is, the percentage of rice exported, consumed as food, saved for seed, wasted and imported. Control variables measure education expenditure (as % of GNI) (*Education*), agriculture value added per worker at constant prices ($Agr.VA$), agriculture value added as a percentage of GDP ($Agr.VA\%$) and cereal yields in kg per hectare ($CerealsY.$). All variables (dependent and independent) are in logarithms. All estimations contain year fixed effects. Driscoll-Kraay standard errors are in parentheses and are corrected for heteroscedasticity, autocorrelation and cross sectional dependence. We use the following conventional symbols for p-values: *** p<0.01, ** p<0.05, * p<0.1.

Table 2: The relation between research focus (relative specialisation index) and revealed demands proxied by rice relative use

VARIABLES	(1) Plant protect.	(2) Yields	(3) Trad. genetics	(4) Genomics	(5) Human cons.	(6) Prod. practices
Rev. demands						
Rice calories	0.40	-0.40	-0.54***	-0.70**	-0.74***	0.89***
	(0.26)	(0.33)	(0.19)	(0.34)	(0.18)	(0.16)
Rice yields	0.45	-1.32***	-0.61***	-0.24	-0.54	-0.98***
	(0.31)	(0.42)	(0.17)	(0.41)	(0.53)	(0.32)
Rice area	-0.46**	0.32	0.22	0.91**	0.63**	-0.20
	(0.19)	(0.30)	(0.25)	(0.37)	(0.29)	(0.12)
Fertilisers	-0.20**	0.03	0.02	-0.16	-0.09	0.23***
	(0.08)	(0.09)	(0.06)	(0.18)	(0.06)	(0.06)
Controls						
Education	-0.11	-0.29***	-0.02	0.07	0.13	0.25***
	(0.10)	(0.09)	(0.07)	(0.18)	(0.18)	(0.08)
Agr. VA	-0.03	0.17*	-0.31**	-0.35*	0.38***	-0.07
	(0.18)	(0.09)	(0.12)	(0.19)	(0.10)	(0.09)
Agr. VA %	0.19	-0.48**	-0.15	-0.52**	0.12	-0.06
	(0.18)	(0.19)	(0.18)	(0.20)	(0.11)	(0.16)
Cereals Y.	0.07	2.32**	1.63***	0.99	1.03	0.46
	(0.31)	(0.88)	(0.34)	(0.85)	(0.84)	(0.47)
Constant	-5.58	-3.69	-1.62	0.16	-3.11	0.96
	(3.67)	(3.27)	(1.85)	(4.45)	(1.98)	(1.96)
Observations	413	410	413	413	413	413
Number of groups	16	16	16	16	16	16
within R^2	0.193	0.267	0.177	0.153	0.211	0.252
F	2630	3949	2953	3777	6018	239
Prob > F	0	0	0	0	0	0
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Notes: results of a fixed effects panel data regression. The dependent variables in the different columns measure the relative specialisation index (in one period/country) for publications in one of the six topics identified in Figure 1 and discussed in Section 5.1. All measures of relative specialisation are computed as three year averages. Revealed demand variables measure per capita human intake of daily calories from rice (*Ricecalories*), rice yields measured as kg per hectare (*Riceyields*), amount of per capita arable land devoted to rice (*Ricearea*) and amount of chemical fertilisers used per arable land and permanent crop areas (*Fertilisers*). Control variables measure education expenditure (as percentage of GNI) (*Education*), agriculture value added per worker at constant prices (*Agr.VA*), agriculture value added as percentage of GDP (*Agr.VA%*), and cereal yields measured as kg per hectare (*CerealsY.*). All variables (dependent and independent) are in logarithms. All estimations contain year fixed effects. Driscoll-Kraay standard errors are in parentheses and are corrected for heteroscedasticity, autocorrelation and cross sectional dependence. We use the following conventional symbols for p-values: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: The relation between research focus (relative specialisation index) and revealed demands proxied by potential needs

(a)	(b)	(c)
Variable	**Expected topic**	**Associated topic**
export (H)	! Human consumption; yields; trad. genetics; genomics	Plant protection (+); Production practices (-)
food (H)	Human consumption; trad. genetics; genomics	Production practices (+); genomics (-)
seeds (H)	Traditional genetics	Trad. genetics (+); human consumption (+); production practices (-)
waste (H)	Production practices; human consumption	Trad. genetics (+); human consumption (-)
imports (H)	Yields; plant protection; traditional genetics; human consumption	Production practices (-)
processed (H)	Human consumption	Human consumption (+)
Ricecalories (H)	Human consumption	Trad. genetics (-); genomics (-); human consumption (-); production practices (+)
Riceyields (H)	! Yields	Yields (-); trad. genetics (-); production practices (-)
Fertilisers (H)	Yields	Plant protection (-); production practices (+)
Ricearea (H)	All	Plant protection (-); trad. genetics; genomics
Undernourished (H)	Human consumption	Genomics (+); plant protection (-)
Pesticides (H)	Plant protection	Plant protection (-)

Notes: column (a) reports the variable used to measure a given revealed demand: the rows 1-7 distinguish among different uses of rice in a country (together with *other* the sum of the different percentage is 1), and the remaining rows are further indicators of revealed demands related to needs; column (b) lists the topics that we expect to be most related to societal demand: in some cases more than one, in other cases, in the same row, we indicate with ! topic(s) that we do not expect to be related to societal demand – some of the research topics are not listed here because they could be relevant to multiple societal demands and not one or more specific demands, e.g., genomics; column (c) lists the topics that were found to be significantly related to the societal demand – '+' is used when the association is positive and '-' is used when the association is negative. Topics are computed from the publication corpus and are described in Section 5.1. All results are discussed in Section 5.2.

Table 4: Summary of results: expected and estimated associations between research topics and societal demands

B.2 Web Appendix

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Plant protect.	Yields	Trad. genetics	Genomics	Human cons.	Prod. practices
Export %	0.07***	-0.01	-0.00	-0.09	0.01	-0.06***
	(0.02)	(0.02)	(0.03)	(0.06)	(0.03)	(0.02)
Food %	0.05	0.45	-0.37	-1.39*	-0.07	1.08**
	(0.36)	(0.28)	(0.39)	(0.69)	(0.44)	(0.40)
Seed %	-0.22	0.06	0.48***	-0.55	0.27**	-0.24***
	(0.15)	(0.18)	(0.17)	(0.39)	(0.12)	(0.07)
Waste %	-0.25	0.13	0.72***	-0.18	-0.53***	-0.15
	(0.15)	(0.13)	(0.07)	(0.29)	(0.14)	(0.10)
Import %	-0.04	0.02	-0.03	0.02	-0.00	-0.04*
	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)
Controls						
Education	-0.17**	-0.30***	-0.01	0.17	0.05	0.18***
	(0.08)	(0.09)	(0.06)	(0.19)	(0.18)	(0.06)
Agr. VA	-0.20	0.29*	-0.25	0.07	0.38*	-0.07
	(0.14)	(0.15)	(0.19)	(0.37)	(0.20)	(0.10)
Agr. VA %	0.15	-0.11	-0.48**	-0.11	0.69***	0.14
	(0.20)	(0.17)	(0.19)	(0.31)	(0.15)	(0.17)
Cereals Y.	0.47*	0.17	0.65***	-0.59	0.05	-0.54***
	(0.26)	(0.17)	(0.23)	(0.48)	(0.23)	(0.19)
Constant	-6.16	-4.11*	-0.50	-1.24	-7.14***	1.49
	(4.00)	(2.33)	(2.54)	(4.81)	(2.23)	(2.44)
Observations	303	303	303	303	303	303
Number of groups	11	11	11	11	11	11
within R^2	0.814	0.373	0.416	0.404	0.633	0.323
F	28988	5795	297.8	984.8	13540	109
Prob > F	0	0	0	0	0	0
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Notes: results of a fixed effects panel data regression. The dependent variables in the different columns measure the relative percentage (in one period/country) of publications in one of the six topics identified in Figure 1 and discussed in Section 5.1. All measures of relative specialisation are computed as three year averages. All variables (dependent and independent) are in logarithms. All estimations include year fixed effects. Driscoll-Kraay standard errors are in parentheses and are corrected for heteroscedasticity, autocorrelation and cross sectional dependence. We use the following conventional symbols for p-values: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: The relation between research focus (percentage) and revealed demand proxied by rice relative use

VARIABLES	(1) Plant protect.	(2) Yields	(3) Trad. genetics	(4) Genomics	(5) Human cons.	(6) Prod. practices
Rice calories	0.40	-0.41	-0.54***	-0.70**	-0.72***	0.90***
	(0.26)	(0.32)	(0.18)	(0.33)	(0.18)	(0.16)
Rice yields	0.45	-1.32***	-0.59***	-0.22	-0.52	-0.96***
	(0.31)	(0.42)	(0.18)	(0.42)	(0.52)	(0.33)
Rice area	-0.46**	0.33	0.23	0.90**	0.62**	-0.21*
	(0.19)	(0.30)	(0.24)	(0.37)	(0.29)	(0.12)
Fertilisers	-0.20**	0.02	0.02	-0.16	-0.08	0.23***
	(0.08)	(0.09)	(0.06)	(0.17)	(0.06)	(0.06)
Controls						
Education	-0.11	-0.29***	-0.02	0.07	0.13	0.26***
	(0.10)	(0.10)	(0.07)	(0.18)	(0.18)	(0.08)
Agr. VA	-0.03	0.17*	-0.32***	-0.35*	0.39***	-0.07
	(0.18)	(0.09)	(0.11)	(0.19)	(0.10)	(0.09)
Agr. VA %	0.19	-0.49**	-0.16	-0.52**	0.12	-0.05
	(0.18)	(0.19)	(0.18)	(0.20)	(0.11)	(0.16)
Cereals Y.	0.08	2.34**	1.62***	0.96	0.99	0.42
	(0.31)	(0.90)	(0.34)	(0.83)	(0.82)	(0.49)
Constant	-7.81**	-5.67*	-3.57*	-1.45	-4.76**	-0.72
	(3.73)	(3.33)	(1.77)	(4.50)	(1.97)	(1.93)
Observations	413	410	413	413	413	413
Number of groups	16	16	16	16	16	16
within R^2	0.692	0.330	0.229	0.456	0.552	0.187
F	34413	28874	8904	100144	23451	10242
Prob > F	0	0	0	0	0	0
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Notes: results of a fixed effects panel data regression. The dependent variables in the different columns measure the relative percentage (in one period/country) of publications in one of the six topics identified in Figure 1 and discussed in Section 5.1. All measures of relative specialisation are computed as three year averages. Revealed demand variables measure per capita human intake of daily calories from rice (*Ricecalories*), rice yields measured as kg per hectare (*Riceyields*), amount of per capita arable land devoted to rice (*Ricearea*), and amount of chemical fertilisers used per arable land and permanent crop areas (*Fertilisers*). Control variables measure education expenditure (as a percentage of GNI) (*Education*), agriculture value added per worker at constant prices (*Agr.VA*), agriculture value added as a percentage of GDP (*Agr.VA%*), and cereal yields measured as kg per hectare (*CerealsY.*). All variables (dependent and independent) are in logarithms. All estimations include year fixed effects. Driscoll-Kraay standard errors are in parentheses and are corrected for heteroscedasticity, autocorrelation and cross sectional dependence. We use the following conventional symbols for p-values: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: The relation between research focus (percentage) and revealed demands proxied by potential needs

VARIABLES	(1) Plant protect.	(2) Yields	(3) Trad. genetics	(4) Genomics	(5) Human cons.	(6) Prod. practices
Export %	0.07***	-0.02	0.03	-0.08	0.02	-0.05**
	(0.01)	(0.02)	(0.03)	(0.06)	(0.03)	(0.02)
Food%	0.13	0.31	0.03	-1.14	-0.41	1.07***
	(0.27)	(0.30)	(0.38)	(0.72)	(0.40)	(0.32)
Seed %	-0.17	-0.00	0.52***	-0.23	0.25**	-0.18**
	(0.15)	(0.16)	(0.14)	(0.28)	(0.10)	(0.06)
Waste %	-0.19	0.06	0.77***	-0.14	-0.60***	-0.15
	(0.12)	(0.11)	(0.09)	(0.25)	(0.12)	(0.10)
Import %	-0.06**	0.00	0.01	0.01	0.00	-0.00
	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)
Education	-0.21**	-0.27**	0.05	0.10	0.04	0.19**
	(0.07)	(0.09)	(0.06)	(0.16)	(0.17)	(0.07)
Agr. VA	-0.23*	0.32**	-0.37*	-0.01	0.44***	-0.05
	(0.11)	(0.14)	(0.17)	(0.34)	(0.13)	(0.08)
Agr. VA %	0.10	-0.13	-0.40**	-0.08	0.65***	0.11
	(0.15)	(0.19)	(0.14)	(0.28)	(0.15)	(0.16)
Cereals Y.	0.31	0.17	0.73***	-0.57	0.07	-0.37**
	(0.17)	(0.14)	(0.21)	(0.40)	(0.20)	(0.15)
Constant	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	294	294	294	294	294	294
Number of groups	11	11	11	11	11	11
within R^2	0.386	0.264	0.384	0.110	0.316	0.449
F	165.4	12.19	742	219.6	115.7	102.2
Prob > F	3.28e-10	0.000102	0	8.03e-11	1.94e-09	3.58e-09
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Notes: results of a fixed effects panel data regression. The dependent variables in the different columns measure the relative percentage (in one period/country) of publications in one of the six topics identified in Figure 1 and discussed in Section 5.1. All measures of relative specialisation are computed as a three years average. All variables (dependent and independent) are in logarithms. *Export*, *Food*, *Seed*, *Waste*, *Import* are all lagged by one year. All estimations include year fixed effects. Driscoll-Kraay standard errors are in parentheses and are corrected for heteroscedasticity, autocorrelation and cross sectional dependence. We use the following conventional symbols for p-values: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: The relation between research focus (relative specialisation index) and revealed demands from rice relative use in $t-1$

VARIABLES	(1) Plant protect.	(2) Yields	(3) Trad. genetics	(4) Genomics	(5) Human cons.	(6) Prod. practices
Rice calories	0.27	-0.39	-0.52***	-0.51	-0.75***	0.97***
	(0.17)	(0.29)	(0.10)	(0.31)	(0.19)	(0.18)
Rice yields	0.75***	-0.92**	-0.54**	-0.28	-0.30	-0.99***
	(0.22)	(0.32)	(0.20)	(0.38)	(0.33)	(0.22)
Rice area	-0.27	0.29	0.09	0.85*	0.62**	-0.25
	(0.20)	(0.25)	(0.22)	(0.41)	(0.25)	(0.16)
Fertilisers	-0.21***	0.02	0.02	-0.04	-0.10*	0.25***
	(0.07)	(0.09)	(0.07)	(0.16)	(0.05)	(0.06)
Education	-0.14	-0.29***	0.00	0.04	0.11	0.27***
	(0.09)	(0.08)	(0.06)	(0.15)	(0.17)	(0.07)
Agr. VA	-0.04	0.16	-0.33***	-0.43**	0.34***	-0.12
	(0.13)	(0.11)	(0.09)	(0.16)	(0.09)	(0.07)
Agr. VA %	0.12	-0.41*	-0.14	-0.53***	0.11	-0.00
	(0.13)	(0.20)	(0.18)	(0.16)	(0.12)	(0.13)
Cereals Y.	-0.15	1.93**	1.44***	0.51	0.68	0.44
	(0.20)	(0.77)	(0.32)	(0.55)	(0.59)	(0.37)
Constant	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	403	401	403	403	403	403
Number of groups	16	16	16	16	16	16
within R^2	0.191	0.246	0.173	0.163	0.210	0.283
F	10607	10896	2442	884	4063	449.4
Prob > F	0	0	0	0	0	0
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Notes: results of a fixed effects panel data regression. The dependent variables in the different columns measure the relative percentage (in one period/country) of publications in one of the six topics identified in Figure 1 and discussed in Section 5.1. All measures of relative specialisation are computed as three year averages. Revealed demand variables measure per capita human intake of daily calories from rice ($Ricecalories$), rice yields measured as kg per hectare ($Riceyields$), amount of arable land per capita devoted to rice ($Ricearea$), and amount of chemical fertilisers used per arable land and permanent crop areas ($Fertilisers$). Control variables measure education expenditure (as a percentage of GNI) ($Education$), agriculture value added per worker at constant prices ($Agr.VA$), agriculture value added as a percentage of GDP ($Agr.VA\%$), and cereal yields measured as kg per hectare ($CerealsY.$). All variables (dependent and independent) are in logarithms. Variables *Rice calories*, *Rice yields*, *Rice area*, and *Fertilisers* are lagged by one time period. All estimations include year fixed effects. Driscoll-Kraay standard errors are in parentheses and are corrected for heteroscedasticity, autocorrelation and cross sectional dependence. We use the following conventional symbols for p-values: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: The relation between research focus (relative specialisation index) and revealed demands from rice relative needs in $t-1$

VARIABLES	(1) Plant protect.	(2) Yields	(3) Trad. genetics	(4) Genomics	(5) Human cons.	(6) Prod. practices
Export %	0.06***	-0.04*	0.05	-0.03	0.01	-0.04*
	(0.01)	(0.02)	(0.03)	(0.05)	(0.03)	(0.02)
Food %	0.21	0.22	0.13	-0.91	-0.74	1.07***
	(0.26)	(0.25)	(0.34)	(0.76)	(0.45)	(0.28)
Seed%	-0.06	-0.17	0.52***	-0.04	0.19*	-0.05
	(0.12)	(0.13)	(0.15)	(0.27)	(0.10)	(0.06)
Waste %	-0.14	-0.04	0.71***	-0.06	-0.69***	-0.14
	(0.10)	(0.10)	(0.11)	(0.23)	(0.12)	(0.12)
Import %	-0.09***	0.00	0.04	0.03	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Education	-0.23***	-0.25**	0.13**	0.08	0.01	0.18**
	(0.07)	(0.09)	(0.06)	(0.15)	(0.17)	(0.07)
Agr. VA	-0.24**	0.36**	-0.43**	-0.11	0.49***	-0.09
	(0.09)	(0.14)	(0.16)	(0.38)	(0.10)	(0.08)
Agr. VA %	0.04	-0.14	-0.39**	-0.07	0.65***	0.05
	(0.13)	(0.18)	(0.13)	(0.26)	(0.14)	(0.14)
Cereals Y.	0.25	0.17	0.87***	-0.52	0.06	-0.39**
	(0.15)	(0.15)	(0.24)	(0.41)	(0.26)	(0.16)
Constant	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	283	283	283	283	283	283
Number of groups	11	11	11	11	11	11
within R^2	0.412	0.257	0.386	0.0980	0.350	0.430
F	157.8	14.35	1037	95.21	328.4	45.75
Prob > F	4.14e-10	4.81e-05	0	5.08e-09	0	1.87e-07
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Notes: results of a fixed effects panel data regression. The dependent variables in the different columns measure the relative percentage (in one period/country) of publications in one of the six topics identified in Figure 1 and discussed in Section 5.1. All measures of relative specialisation are computed as three year averages. All variables (dependent and independent) are in logarithms. *Export*, *Food*, *Seed*, *Waste*, *Import* and are lagged by two years. All estimations include year fixed effects. Driscoll-Kraay standard errors are in parentheses and are corrected for heteroscedasticity, autocorrelation and cross sectional dependence. We use the following conventional symbols for p-values: *** p<0.01, ** p<0.05, * p<0.1.

Table 9: The relation between research focus (relative specialisation index) and revealed demands from rice relative use in $t-2$

VARIABLES	(1) Plant protect.	(2) Yields	(3) Trad. genetics	(4) Genomics	(5) Human cons.	(6) Prod. practices
Rice calories	0.18	-0.19	-0.47***	-0.35	-0.85***	1.00***
	(0.14)	(0.31)	(0.10)	(0.26)	(0.20)	(0.20)
Rice yields	0.83***	-0.70**	-0.69***	-0.30	0.13	-0.93***
	(0.25)	(0.27)	(0.21)	(0.50)	(0.26)	(0.16)
Rice area	-0.16	0.15	-0.07	0.87*	0.69**	-0.20
	(0.27)	(0.26)	(0.22)	(0.48)	(0.25)	(0.21)
Fertilisers	-0.21***	0.02	0.06	0.07	-0.18***	0.28***
	(0.06)	(0.08)	(0.07)	(0.17)	(0.04)	(0.06)
Education	-0.16*	-0.26***	0.03	0.01	0.08	0.27***
	(0.09)	(0.08)	(0.07)	(0.15)	(0.17)	(0.07)
Agr. VA	-0.02	0.18	-0.36***	-0.46***	0.30**	-0.15**
	(0.12)	(0.12)	(0.09)	(0.15)	(0.11)	(0.06)
Agr. VA %	0.03	-0.36	-0.09	-0.55***	0.15	-0.01
	(0.13)	(0.21)	(0.19)	(0.16)	(0.12)	(0.12)
Cereals Y.	-0.15	1.66**	1.41***	0.26	0.21	0.34
	(0.15)	(0.73)	(0.33)	(0.49)	(0.41)	(0.33)
Constant	-7.10	0.00	1.66	4.79	-1.91	1.08
	(4.33)	(0.00)	(2.28)	(5.89)	(2.67)	(1.55)
Observations	388	387	388	388	388	388
Number of groups	16	16	16	16	16	16
within R^2	0.191	0.221	0.190	0.176	0.221	0.291
F	449.9	64715	1613	1334	6338	1486
Prob > F	0	0	0	0	0	0
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes

Notes: results of a fixed effects panel data regression. The dependent variables in the different columns measure the relative percentage (in one period/country) of publications in one of the six topics identified in Figure 1 and discussed in Section 5.1. All measures of relative specialisation are computed as three year averages. Revealed demand variables measure the per capita human intake of daily calories from rice (*Ricecalories*), rice yields measured as kg per hectare (*Riceyields*), amount of per capita arable land devoted to rice (*Ricearea*), and amount of chemical fertilisers used per arable land and permanent crop areas (*Fertilisers*). Control variables measure the education expenditure (as a percentage of GNI) (*Education*), agriculture value added per worker at constant prices ($Agr.VA$), agriculture value added as a percentage of GDP ($Agr.VA\%$), and cereal yields measured as kg per hectare ($CerealsY.$). All variables (dependent and independent) are in logarithms. Variables *Rice calories*, *Rice yields*, *Rice area* and *Fertilisers* are lagged by two time periods. All estimations include year fixed effects. Driscoll-Kraay standard errors are in parentheses and are corrected for heteroscedasticity, autocorrelation and cross sectional dependence. We use the following conventional symbols for p-values: *** p<0.01, ** p<0.05, * p<0.1.

Table 10: The relation between research focus (relative specialisation index) and revealed demands from rice relative needs in $t-2$

C Assessing First Author Affiliation Potential Bias

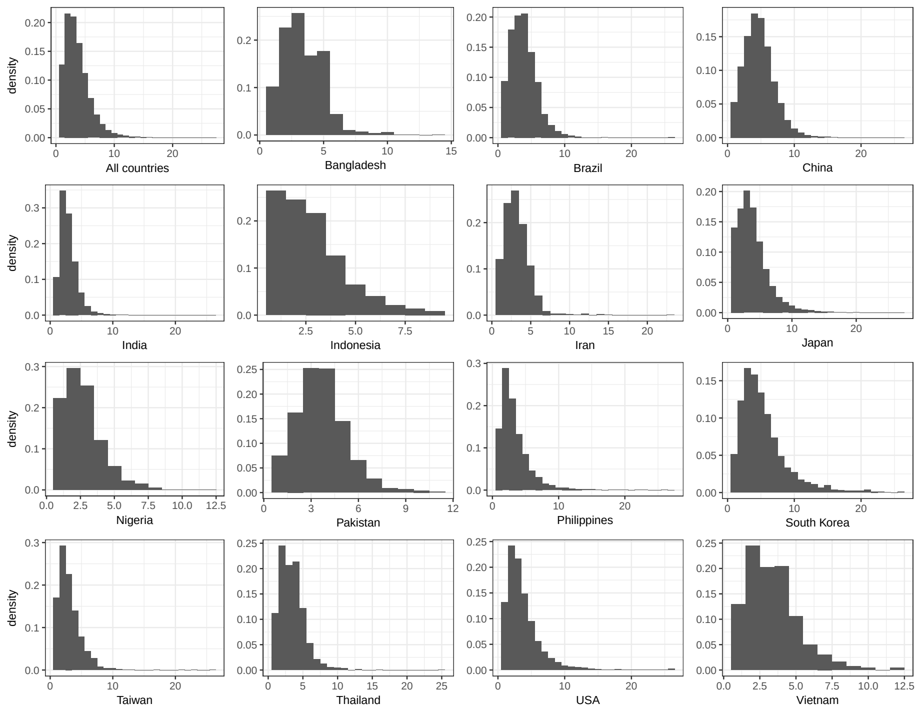
In this section we discuss the extent to which attributing a publication to a country based on first author's affiliation (the only information available in CAB abstract) may affect our results. The answer is that it does not, as we explain below.

Our aim is to uncover the share of a country's publications devoted to a given topic, rather than to evaluate a country's publication performance. Therefore, we are interested in the attribution of a topic to a country, to measure its relative relevance, rather than in a country's number of publications. Therefore, in our case, a problem would arise should there be systematic distortion in the assignment of authorship across different countries when using the first author affiliation. First, let us consider the case where all papers are single authored: this would not cause a problem. Then, let us consider the case where publications are shared between single and dual authored papers. An issue might arise if, in country A, all publications are single authored and in country B, all publications are dual authored. This would mean that, for country A, first author's affiliation would provide information on where the research was conducted; however, this may not apply to country B (if some of the second authors are from a different country). This would introduce a systematic difference between country A and B, implying that allocation of first author affiliation may not be random (which is a desirable property to obtain unbiased results).

We first check the numbers of authors across countries. On average, across countries, publications have just less than three authors (median 3). Figure 10 plots the distribution of publications per number of authors, for most of the countries in the estimated sample. Most countries follow a quite similar highly skewed distribution, with most publications written by three or four authors and a small share of publications written by a large number of authors. This suggests that they all have a similar probability of having one author from a different country. However, while countries such as Brazil, Japan, Philippines, Thailand, and the USA are closest to the overall distribution (top left panel), other countries differ – e.g. Bangladesh, Indonesia and Pakistan present a less skewed distribution.

Figure 11 plots the share of publications with a given number of authors for each of the countries in the estimated sample, from 1 (bottom dark shaded interval) to 9 (top light shaded interval). It shows that all countries follow a similar publications pattern, although with some international differences with respect to co-authorship. However, it is not clear whether those differences imply differences in the number of international collaborations. For instance, the patterns for Japan, Iran and the USA are similar, although international collaborations in the three countries are likely to differ; Cuba is more similar to Brazil than to China, which is instead most similar to South Korea.

In sum, although we do not observe a clear pattern, both plots suggest that there are some differences in publication strategies across countries. However, these differences might be related to the dominant specialisation, reflecting disciplinary differences in co-authorship. For example, India has a substantially stronger focus than China on social



Notes: each panel plots a country's distribution of publications for a given number of authors (between 1983-20120, where the minimum is 1 and the maximum is 27. On the vertical axis we measure the density of publications per bin

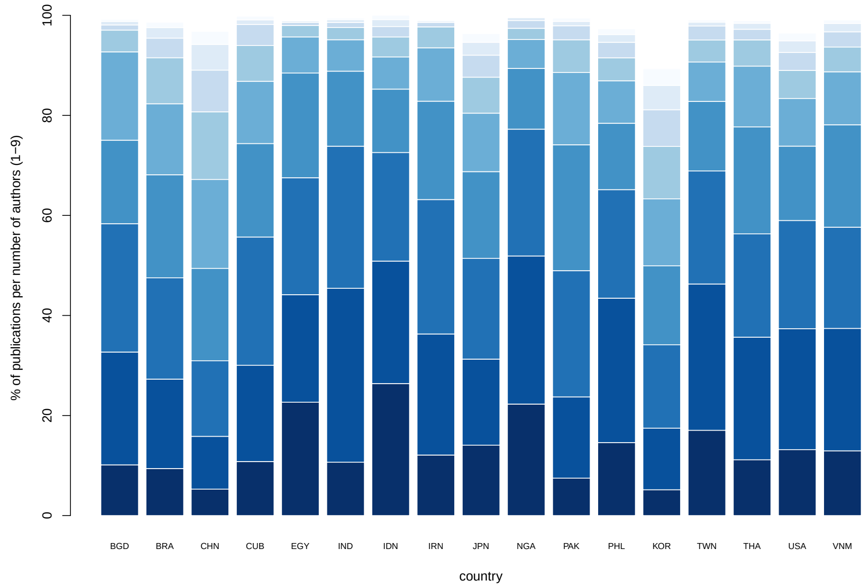
Figure 10: Distribution of publications per number of authors.

sciences and production practices, while China has a stronger focus on genomics.

To verify that these differences in co-authorship patterns do not introduce a bias when only the first author affiliation is available, we exploit the CABI descriptor, which reports the country at the core of the research described in the publication. Unfortunately, this information is available for just under half of the publications in our database. The country CABI descriptor does not necessarily refer to the country on which the research is focused, but is used to index relevant references to geographical concepts in the publication. That is, if there is a country descriptor, this means that the country must be mentioned in the publication for some reason. If the geographical descriptor is missing, this might mean that the information was not included (error) or that the publication does not refer to a specific country.[32]

Figure 12 plots the distribution of publications per number of countries, as reported by the CABI descriptor (1 bin per country). It shows that 96% of the publications mention only one relevant country (although the maximum number of countries mentioned is 93). According to the CABI descriptor, nearly all publications reporting a country study, focus on a single country. This would suggest that when there are multiple authors, and the publication make explicit reference to a specific country, it is unlikely that a the publication is relevant for more than one country. The relevant question is then whether the country

[32]This might be because an author who is from country A and is working on a region r in that country , might not need to refer explicitly to that country in the publication.



Notes: each bar represents a country, labelled using the three digit international codes. The vertical axis measures the share of publications per interval. Each interval represents a number of authors, starting from 1 (dark shaded lowest interval) to 9 (the lightest shaded top interval)

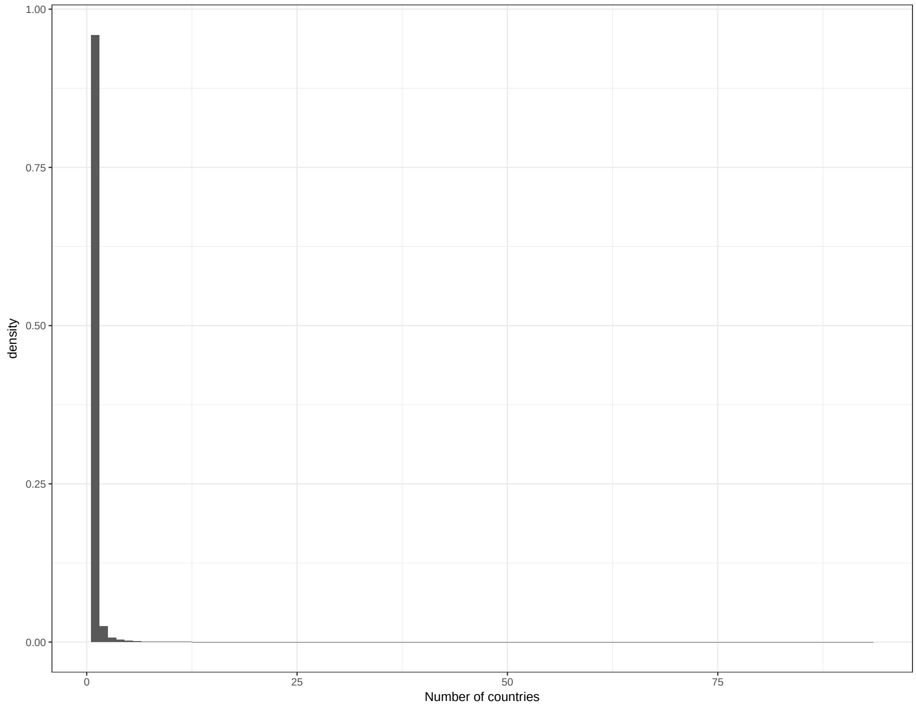
Figure 11: Publications per number of authors

of the first author is the same of the country of study, or if they differ.

We compare the country derived from the first author's affiliation to the country being studied according to the CABI descriptor. When the study is focused on a single country, in 91% of cases, the information coincides: that is, for 91% of the publications mentioning a specific country (and included as CABI descriptor), the country of affiliation of the first author is the same as the country indicated in the publication. The evidence seems to suggest that, if there are differences in publication practices related to the number of authors per publication, the number of authors does not seem to be correlated to the number of countries in which the research is conducted and, therefore, use of the first author's affiliation does not seem to introduce a relevant bias.

To further check this, we computed the share of publications per country in four different scenarios: the whole database (*All*); when the first author affiliation and the CABI geographic descriptor differ (*Diff*); when the first author affiliation and the CABI geographic descriptor coincide (*Same*); and when there is no information on the geographic CABI descriptor (*Missing*). If the share of publications in these four subsets of the database differ, this may be an indication that using first author information is introducing a bias.

We estimate the correlation in countries' publications shares across all countries (including those with very small numbers of publications, which may differ substantially across the four subsets). Table 11 reports the results of the non-parametric Spearman correlation coefficients, which are all significant at the 99% level of confidence. The results suggest that the distribution of publications across countries in the overall database



Notes: distribution of the number of publications per bin, where each bin is represented by the number of countries studied according to the CABI descriptor (from 1 to 93). The vertical axis measures the density of publications for each bin

Figure 12: Publications per number of countries as indicated by the CABI descriptor

(*All*) is highly and significantly correlated to the other subsets, both when information on affiliation and information in the CABI descriptor differ (*Diff*) and when it is the same (*Same*). If anything, the correlation is slightly higher when the information between the two sources differs, suggesting that it may not be use of first affiliation that is introducing a bias with respect to the country for which the research is relevant.

VARIABLES	(1) All	(2) Diff	(3) Same	(4) Missing
All	1.00	0.89	0.87	0.95
Diff	0.89	1.00	0.68	0.87
Same	0.87	0.68	1.00	0.93
Missing	0.95	0.87	0.93	1.00

Notes: .

Table 11: Correlation between countries' publication shares for different subsets of the database

Recent papers in the SPRU Working Paper Series:

April

Knowledge Sources and Impacts on Subsequent Inventions: Do Green Technologies Differ from Non-Green Ones? Nicolò Barbieri, Alberto Marzucchi and Ugo Rizzo

Do Financial Constraints Hamper Environmental Innovation Diffusion? An Agent-Based Approach. Paola D'Orazio and Marco Valente

March

Energy Security, Employment and the Policy-Industry Interlock: Explaining the Role of Multi-Scalar Socio-Spatial Embeddedness in Industry Destabilization. Silver Sillak and Laur Kanger

The Effect of R&D Growth on Employment and Self-Employment in Local Labour Markets. Tommaso Ciarli, Alberto Marzucchi, Edgar Salgado and Maria Savona

Industrial Policy for a European Industrial Renaissance. A Few Reflections. Maria Savona

January

Patent-based Estimation Procedure of Private R&D: The Case of Climate Change and Mitigation Technologies in Europe. Francesco Pasimeni, Alessandro Fiorini and Aliki Georgakaki

Reorienting Finance Towards Energy Efficiency: The Case of UK Housing. Noam Bergman and Tim Foxon

Innovation for Inclusive Structural Change. A Framework and Research Agenda. Tommaso Ciarli, Maria Savona, Jodie Thorpe and Seife Ayele

System Transition and Structural Change Processes in the Energy Efficiency of Residential Sector: Evidence from EU Countries. Valeria Costantini, Francesco Crespi, Elena Paglialunga and Giorgia Sforna

Technological Innovation, Entrepreneurship and Productivity in Germany, 1871-2015. Wim Naudé and Paula Nagler

Suggested citation:
Tommaso Ciarli and Ismael Ràfols (2018). *The Relation between Research Priorities and Societal Demands: The Case of Rice*. SPRU Working Paper Series (SWPS), 2018-12: 1-64. ISSN 2057-6668. Available at: www.sussex.ac.uk/spru/swps2018-12

SPRU – Science Policy Research Unit
University of Sussex
Falmer, Brighton, BN1 9SL,United Kingdom
SWPS Website: www.sussex.ac.uk/spru/research/swps
SPRU Website: www.sussex.ac.uk/spru
SPRU Twitter: @SPRU