WORKING PAPER

Thought for Food: A New Dataset on Innovation for Agricultural Use

By
Daniel K.N. Johnson and Christopher R. Hughes

Colorado College Working Paper 2009-03
June, 2009

Department of Economics and Business
Colorado College
Colorado Springs, Colorado 80903-3298
www.coloradocollege.edu/dept/EC

Thought for Food:
A New Dataset on Innovation for Agricultural Use

by

Daniel K.N. Johnson
and Christopher R. Hughes

June 8, 2009

Abstract

Agriculture, like many primary and service sectors, is a frequent recipient of innovation intended for its use, even if those innovations originate in industrial sectors. The challenge has been identifying them from patent data, which are recorded for administrative purposes using the International Patent Classification (IPC) system. We reprogram a well-tested tool, the OECD Technology Concordance (OTC), to identify 16 million patents granted between 1975 and 2006 worldwide which have potential application in agriculture. This paper presents the methodology of that dataset's construction, introduces the data via summaries by nation and industrial sector over time, and suggests some potential avenues for future exploration of empirical issues using these data.

The dataset described in this paper was commissioned by the US Department of Agriculture's Economic Research Service (ERS), and benefited from the comments of John King in particular.

Daniel K.N. Johnson is the Gerald L. Schlessman Professor of Economics and Director of the Innovative Minds Program, Colorado College's Department of Economics and Business, Colorado Springs, CO, 80903, USA, djohnson@ColoradoCollege.edu, 719-389-6654. Christopher R. Hughes is a recent alumnus of Colorado College's Department of Economics and Business.

Thought for Food:
A New Dataset on Innovation for Agricultural Use

Introduction

While patent data are now readily available for most nations, those data are still of minimal use for economic analysis due to their mode of presentation. Patents are recorded for administrative purposes using the International Patent Classification (IPC) system, which categorizes inventions by product or process. Instead, most economic data and analyses are interested in the particular sectors of the economy responsible for the invention or its subsequent use.

Research into the impact of agricultural innovation suffers grievously under this data structure. After all, what is an agricultural innovation? Should we limit our attention to patents on living organisms? Surely we should widen our vision to include chemicals and machinery, but then which do we include? This paper uses a re-programmed version of the OECD Technology Concordance (Johnson, 2002) to identify patents applicable to agricultural, and further to quantify the probability that any given patent will be used for an agricultural purpose. We then summarize the data for interested researchers, hopefully as a springboard to empirical studies to follow.

Methodology

The goal of the OECD Technology Concordance (OTC) is one of translation, mapping one classification system into another. In particular, it maps patent product or process categories into the economic sectors responsible for their creation and subsequent use. This translation could be done by reading each patent and making individual assignments, but is extremely laborious. Instead, the OTC (like the original Yale Technology Concordance before it), makes use of Canadian Intellectual Property Office data to make the translation.

Upon application for protection, patents are assigned a product code which helps lawyers and patent examiners in grant and litigation decisions. Unfortunately, while the international standard IPC system has existed since 1976, the patent class system is useful almost exclusively for legal purposes. It is of little use to researchers who wish to combine patent data with other datasets, since the product definitions correspond with no other classification system. For example, the IPC grouping B05 includes all goods or processes involved in “spraying or atomizing in general; applying liquids or other fluent materials to surfaces, in general”, and so will include products and processes from a variety of different industries, from cosmetics atomizers to agricultural pesticide sprayers. In short, B05 might very well include cosmetics atomizers alongside pesticide dispensers.

Fortunately, between 1972 and 1995 the Canadian Intellectual Property Office simultaneously assigned IPC codes along with an industry of manufacture (IOM) and sector of use (SOU) code to each of over 300,000 granted patents. For example, in the IPC of B05 (sprayers and atomizers), a cosmetics atomizer might have an IOM in the glass container industry or metal valve industry, while a pesticide sprayer might have an IOM in the chemical fertilizer or agricultural machinery industry. Sectors of use (SOUs) would also differ, with the cosmetics atomizer used in the personal hygiene or cosmetics sector, and the pesticide sprayer used in field crop sectors. The original Yale Technology Concordance (YTC) utilized tabulated information on all 300,000 patents to determine the probability that a patent with a specific IPC has a particular IOM-SOU combination. Since other nations only report IPC information, those probabilities allow researchers to infer the IOM-SOU details of a patent based purely on the legal-technological details offered by the IPC grouping.

We re-programmed the OTC slightly to consider only the probability that a given patent would be associated with any economic activity in the agricultural or food sectors, including crops, livestock, and horticulture along with the processing of food and beverage products. Our 'output' is a probability measure, since our work is based on the application of probabilistic evidence surrounding each patent's IPC. Indeed, even a thorough reading of each patent would conclude that there is no way of knowing with absolute certainty in which economic sector a given patent will eventually be used.

Despite the fact that the OTC is based on Canadian data, its use will not superimpose the industrial structure of Canadian inventions when applied to patent data for other nations. The probabilities of the OTC indicate a technical relationship between IPC (a product/process definition) and IOM or SOU (an industry definition), but permit enormous flexibility for the data to display the industrial composition of patenting in any nation. For further information on the construction, statistical validity and application of the original YTC, see papers in the Johnson and Evenson (1997) volume, Johnson and Evenson (1999), Johnson and Santaniello (2000), or Johnson (2002).

This process is borrowed completely intact from the original YTC, as originally presented by Kortum and Putnam (1997) and subsequently reprogrammed to include simultaneous IOM-SOU assignments as described in Johnson (2002). After reading all 300,000 patents that have information on IPC, IOM and SOU, probabilities were calculated for each IPC to determine the likelihood of any random patent in that IPC having a particular IOM-SOU combination. Thus, with the introduction of a new sample of patents with IPC data attached, it is straightforward to use the probabilities to predict the number of patents with each IOM-SOU combination. This processing transforms a vector of patent data (patents listed by IPC) into a matrix of interrelated patent data (IOM rows and SOU columns).

For example, consider a sample of only 12 patents, 6 each in IPC categories A01B and A01C (making a vector of patent data 2 rows long). Assume that the OTC has determined the following probabilities for each IPC:

patents in A01B have 0.75 probability of having IOM=1, SOU=1
patents in A01B have 0.10 probability of having IOM=1, SOU=2
patents in A01B have 0.15 probability of having IOM=2, SOU=1
patents in A01C have 0.5 probability of having IOM=1, SOU=2
patents in A01C have 0.25 probability of having IOM=1, SOU=3
patents in A01C have 0.25 probability of having IOM=3, SOU=2

Probabilities within each IPC sum to one, meaning that all patents will be transferred into an IOM-SOU combination. For this example, the results will be

Table 1: Sample Output of OECD Technology Concordance

	SOU 1	SOU 2	SOU 3	Total by IOM
IOM 1	4.5	3.6	1.5	**9.6**
2	0.9	0	0	**0.9**
3	0	1.5	0	**1.5**
Total by SOU	**5.4**	**5.1**	**1.5**	**12**

From this result the researcher can tell not only how important each inventing industry of manufacture (IOM) is relative to the total, and not only how important each sector of use (SOU) is relative to the total, but also how important each interaction between IOM and SOU is relative to the total. Our work is considerably easier here, since we consider only one SOU of importance: all agricultural sectors combined.

One limitation of this statistical assignment is that the accuracy of the results depend upon the nature of the original IOM and SOU decisions made by officials at the Canadian Intellectual Property Office. Since decisions were made by patent examiners trained as experts in their respective fields, the accuracy is undoubtedly very high. However, service sectors were never considered to be possible originating sectors for inventions, so OTC results and all that use the same information (including this study) are limited by that decision. Thus, IOMs will always be primary or secondary activities.

For the results below, we used raw data on granted patents from the European Patent Office's PATSTAT database, updated through September of 2006. We limit our consideration to patent documents granted between 1975 and 2006, omitting documents for which errors or missing data are obvious. We also omit consideration of the following nations for logistical reasons, as their national system of record-keeping does not permit us to identify granted patents from patent applications, and we were fearful of counting each relevant document twice: Argentina, Belgium, Colombia, CR, East Germany, EC, Spain, GE, Guatemala, Israel, Luxembourg, Malawi, Mexico, NI, New Zealand, SG, and Thailand. Due to a change in the Japanese Patent Office's recording codes, we were similarly unable to categorically separate patents from published applications in that nation beyond 1996. As Japan represents a large share of the world's total patenting activity, in world totals for the tables and graphs below, Japanese patents are projected in proportion to annual fluctuation in worldwide publications through the remaining years of the study. This methodology provides 16,183,355 patents across 63 nations for analysis.

Results

This section briefly summarizes the dataset, pointing out key nations and industries of manufacture. Throughout, we restrict our attention to patents with possible application to agriculture, counting them in one of two ways: as a simple count of documents, and as a weighted count of documents where the weight is the probability that a given patent will be used in agriculture.

Summary

The global number of agricultural-use patents has grown fairly steadily over the past thirty years, but the world total hides some volatility nation by nation. Figure 1 shows t he number of patents for some key nations, arranged by the year of application. The precipitous decline after 2001 is mostly an artifice of the data, as we used patents granted before December of 2006, and many applications from the preceding 5 years were not yet granted to make our dataset. Notice the rise of China's importance in the last seven year of the period, regardless of this lag problem (evidence of even greater importance once all applications have time to be granted). Japan has slowly climbed to match the US as the single national leader (although remember that post-1996 Japanese values are projected due to data integrity issues). One very interesting story is the importance of the 'rest of the world', nations which are not among the top five patenting authorities.

<u>Figure 1: Patent Counts by Nation over Time</u>

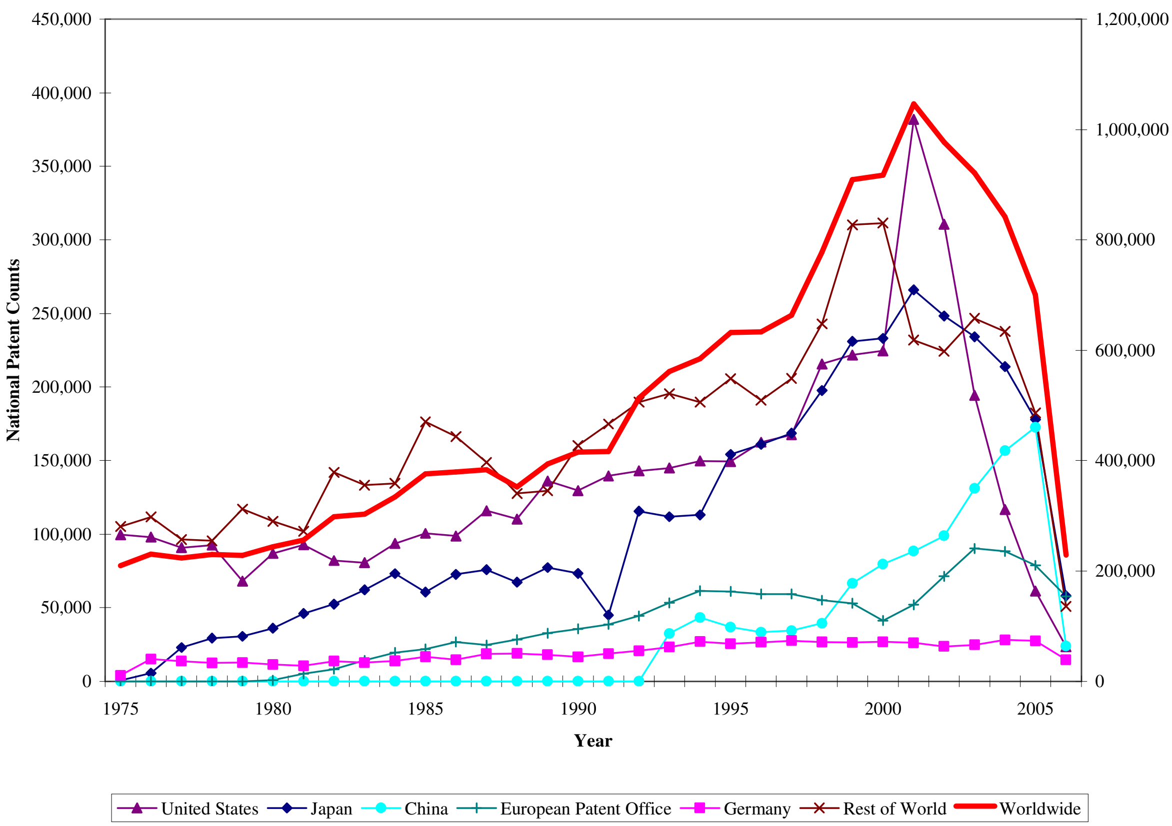

Table 2 summarizes the totals over time for the same five publishing authorities. The United States and Japan were the dominant producers of patents for agricultural use, generating 49 percent of worldwide totals between 1975 and 2006. Since its arrival on the world scene in 1993, China's State Intellectual Property Office (SIPO) has achieved outstanding growth and become a major contributor to agricultural innovation, now claiming a full six percent of the thirty-year total. By comparison, the EPO claims just a slight advantage in raw document counts, with Germany granting a little more than half of that number. Publications from 58 other nations combine to form a single Rest of the World (ROW) total.

Table 2: Patents by Major Authority, 1975-2006

Nation	Unweighted Patent Count	% Unweighted	Weighted Patent Count	% Weighted
United States of America	4,383,801	27%	150,973	23%
Japan	3,515,460	22%	133,359	20%
China	1,038,064	6%	57,197	9%
European Patent Office	1,182,467	7%	45,497	7%
Germany	617,688	4%	38,541	6%
Rest of World	5,445,856	34%	238,826	36%
World Wide	16,183,335	100%	664,392	100%

Nations of the former Soviet Union did not publish enough patents to justify direct comparison to the nations reviewed in this study. However, its dissolution has implications worth considering. In 1993 we observe a 1000 percent increase in publications from the aggregate total of the former Soviet (Goskomizobretenie) & Russian (Rospatent) patent offices. This distortion and sustained high levels by Rospatent suggest that Goskomizobretenie may have underrepresented innovation as measured by patent publications before 1991.

Notice from the weighted columns in Table 2 that the US and Japan fall in importance relative to other patenting authorities. This indicates that while they have a large number of documents granted that have some relevance to agriculture, the average relevance is lower than that of their peer nations (their weights are smaller). China in particular seems to have more heavily agriculture-weighted patents.

Figure 2 compares the industrial origins of weighted patents for agricultural use in the same cohort of patenting authorities. While the raw data provide detail across 37 sectors, we present only six aggregate sectors here. Interestingly, patents from all agricultural sources represented an average of less than three percent of patents subsequently used in agriculture. In short, agricultural innovation does not, for the most part, happen in agriculture, but happens in another sector for subsequent application in agriculture. While the pattern is fairly similar across nations, China looks more like Germany and the US looks more like Japan, the two sets differing primarily in how important chemicals and drugs are relative to instruments and manufacturing.

Table 3 presents the same information, for weighted versus unweighted patents. Among patents which had some agricultural use, those in "other machinery" were on average most likely to be used in agriculture, as evidenced by their high weighted/unweighted ratio. At the other extreme, there are a host of patents in the "Miscellaneous other sectors" category which have some conceivable application to agriculture, but average a very low probability of that usage (a low weighted/unweighted ratio). The highest weighted/unweighted ratio among our 37 sectors, not shown in the table, is unsurprisingly for innovations hailing from the agricultural sector.

Figure 2: Decomposition of industrial origins for patents of agricultural use

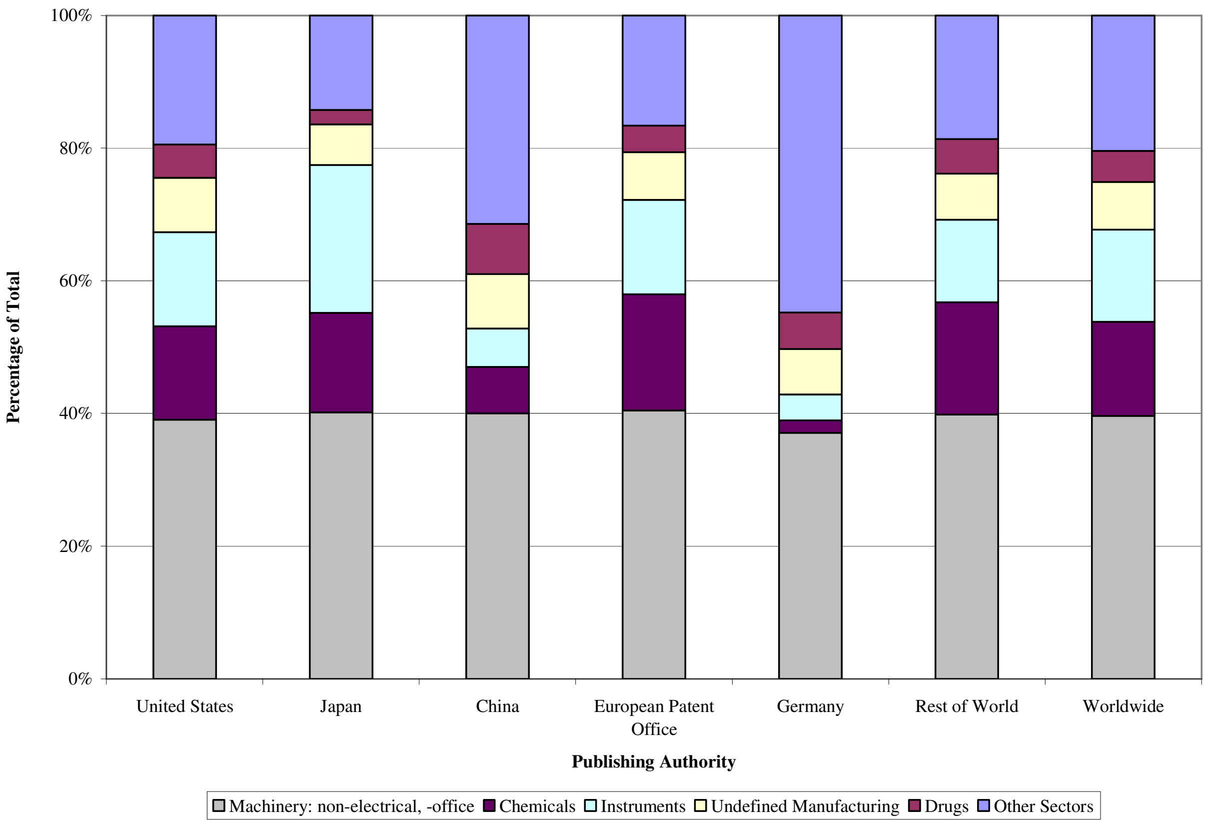

Table 3: Patent Publication by IOM

Sector	Unweighted Patent Count	% Unweighted	Weighted Patent Count	% Weighted
Machinery: non-electrical, -office	3,476,756	21%	263,210	40%
Chemicals	2,350,845	15%	94,397	14%
Instruments	2,129,789	13%	92,152	14%
Undefined Manufacturing	2,801,421	17%	47,798	7%
Drugs	687,534	4%	31,089	5%
Miscellaneous Other Sectors	4,736,990	29%	135,748	20%
Total	16,183,335	100%	664,392	100%

Conclusion

This dataset expands the frontier of agricultural innovation research, enabling several different directions of research to proceed from here. We have merely created the dataset, describing it here for those (including ourselves) who intend to use it for applied purposes.

One direction for future research would be to explain the patterns we see, the sectoral differences between nations and the differences in national emphases on agricultural-use innovation over time. Do our data reflect the result of national policies, or do they reflect different research productivities? How have national policies--- ranging from direct funding of research to stronger intellectual property protection to agricultural extension programs to help diffusion of innovations--- affected different nations over time?

A second interesting research direction would take these data as a starting point to measure knowledge spillovers. How have innovations from diverse industrial sectors contributed to agricultural total factor productivity? Given their costs then, which sectors should research funding target for the greatest productivity impact per dollar?

In short, this dataset of 16 million patents is ripe for econometric research. We encourage interested readers to contact the authors for access to the full set for their own analyses.

References

Evenson, R.E. and D.K.N. Johnson (1997), "Introduction: Invention Input-Output Analysis". Economic Systems Research, Vol 9(2): 149-160.

Harnett, D.L. (1982), Statistical Methods. Addison-Wesley Publishing Company.

Johnson, D.K.N. and V. Santaniello (2000), "Biotechnology Inventions: What Can We Learn From Patents?" Agriculture and Intellectual Property Rights: Economic, Institutional and Implementation Issues in Biotechnology, edited by V. Santaniello et al, CABI Publishing.

Johnson, D.K.N. (2002), The OECD Technology Concordance (OTC): Patents by Industry of Manufacture and Sector of Use, Organization for Economic Cooperation and Development, Science Technology and Industry Working Paper 2002-5.

Johnson, D.K.N. and R.E. Evenson (1999), "R&D Spillovers to Agriculture: Measurement and Applications". Contemporary Economic Policy, 17(4): 432-456.

Johnson, D.K.N. and R.E. Evenson (1997), "Innovation and Invention in Canada". Economic Systems Research, Vol 9(2): 177-192.

Kortum, S. and J. Putnam (1997), "Assigning Patents to Industries: Tests of the Yale Technology Concordance". Economic Systems Research, Vol 9(2): 161-176.

Pray, C.E., K.O. Fuglie and D.K.N. Johnson (2007), “Private Agricultural Research”.Chapter 49 of Handbook of Agricultural Economics Volume 3. North-Holland Elsevier Press.

Statistics Canada. Definitions of Standard Industrial Classification system and Concordance between SIC and ISIC systems. Accessed via US Federal Trade Commission at www.eiit.org/Trade.Resources

World Intellectual Property Office. Definitions of International Patent Classification system. http://www.wipo.org/classifications/en/index.html