How deep are the roots of agricultural innovation? Evidence from patents

Matthew S. Clancy, USDA Economic Research Service, matthew.clancy@ers.usda.gov

Selected Paper prepared for presentation at the 2018 Agricultural & Applied Economics Association Annual Meeting, Washington, D.C., August 5-August 7

How deep are the roots of agricultural innovation? Evidence from patents

*Matthew S. Clancy**

Abstract: We develop a new dataset to study how previously discovered knowledge is used in agricultural R&D. Using patents as instantiations of new commercially valuable agricultural inputs, we interpret the citations patents make as signals of knowledge flows in six different agricultural subsectors over 1976-2016.

We document a number of stylized facts. Agricultural patents tend to cite older ideas than non-agricultural patents, primarily due to the persistence of older ideas. This trend has become more pronounced over the last few decades. Presented another way, the ideas generated in the 1970s and 1980s are more heavily cited, after controlling for various trends in patenting, than ideas generated in the 2000s and 2010s. We obtain similar findings at the level of the individual patent. Agricultural patents are more valuable and more likely to be cited if the age distribution of their ideas has a low mean but a high variance (meaning a long tail of older ideas). We also document variation across sectors: hardware-based agricultural R&D, rather than biochemical agricultural R&D, most heavily relies on older ideas.

* Matt Clancy (matthew.clancy@ers.usda.gov) is an economist at the USDA Economic Research Service. The views expressed herein do not necessarily represent the views of the USDA or the Economic Research Service.

New agricultural technologies are the fruits of research and development (R&D). A considerable literature documents the positive effect of R&D, both private and public, on agricultural productivity (e.g., Huffman and Evenson, 1993, Chavas and Cox 1992, Huffman and Evenson 2006, Alston et al. 2010). However, a major challenge this literature faces is determining the lag between when R&D is conducted and when it begins to impact farm-level productivity. Clearly, it takes time for new discoveries to be understood, verified, developed into agricultural inputs, and for those inputs to be adopted by farmers. On the other hand, after a long time, discoveries may become irrelevant to current technological needs. Just how deep are the roots of agricultural innovations?

The answer chosen can have a significant impact on the estimated rate of return of agricultural R&D. In this literature, R&D is generally thought of as a long-lived capital investment that accrues as firms conduct R&D, but also depreciates as knowledge grows obsolete and applications are exhausted. This capital investment is usually constructed from the sum of lagged R&D expenses, where different lags are assigned varying weights. There have been a variety of methods used to derive the weights attached to different years of R&D: key stages of technological innovation have been used to motivate assumed lag weights; a variety of different assumptions have been subjected to goodness-of-fit tests; or uncertainty has been explicitly modeled in a Bayesian context (Fuglie et al. 2017). In all cases, however, the contribution of R&D to productivity is low in initial years, rises to peak contribution as the R&D is translated into technologies adopted on the farm, and subsequently falls to zero as the knowledge grows less applicable.

In this paper, I use patent data on six agricultural subsectors (ag machinery, animal health products, biocides, biofuel products and processes, plant cultivars, and “other ag”) to provide new evidence on the lag between knowledge production and commercial application in agriculture. Patents can be interpreted as a realization of a successful innovation with commercial value. Patent documents include citations to related literature and these citations are a noisy indicator of knowledge flows. Jaffe, Trajtenberg and Fogarty (2000) find citations represent “clear knowledge flows” or “possible knowledge flows” in approximately 50% of citations to patents. However, even when citations do not indicate direct knowledge flows (e.g., because the citer did not know of the existence of the patented invention), they still signal technological similarity. This is itself a potential indicator that the patents draw on the (unobserved) outputs of similar research programs.

The use of patents as an indicator of innovation has a long history in empirical economics (Nagaoka, Motohashi, and Goto 2010), but has so far found limited use to study agricultural innovation, with a few exceptions (Albers et al. 2016, Pardey et al. 2013). This is partly due to the fact that, historically, most R&D in agriculture was conducted by the public sector (Huffman and Evenson 2006), and the nature of this R&D in the past tended to leave few traces of research outputs in the form of patents. This has changed in the last few decades, due to a combination of factors: the general broadening of the scope of patents to encompassed biological innovations; the 1980 Bayh-Dole act promoting the use of patents by public institutions; and, a drastically increased role for private industry in agricultural R&D (Pray and Fuglie 2015; Clancy and

Moschini 2017). As a result, patents now cover a large and increasing share of US agricultural innovations.

I use the year patent references were published as a new source of evidence on the length of the R&D lag. My patent datasets are chosen to primarily reflect the patents associated with farm-level inputs, rather than agricultural *research* inputs. They are intended to provide evidence on the time between when new discoveries are publicized and when they are translated into commercial products. It is important to note this is *not* the same as the time between R&D spending and farm-level impacts on TFP. This is for two reasons. First, on the citation side, the year a citation is published is not the same as the year research spending was undertaken because research takes time to do. Moreover, patent applications take time to review and non-patent references such as journal articles take time to go through peer review. Second, the year a patent for an agricultural product is granted is not the same year as the year it fully impacts farm-level productivity. Products may or may not be adopted by farmers after a patent is granted, or in some cases, before.

That said, this paper still makes several contributions. First, our data on research lags in patents can be interpreted as providing evidence on the minimum research lags in agriculture. A more accurate measure would need to make assumptions about the speed of diffusion of patented technologies, and the time needed to translate R&D spending into research outputs (both patents and non-patent). This is beyond the scope of our paper, but even our minimum estimates are informative and generally supportive of longer estimates of R&D lags.

Second, we provide novel evidence on trends in R&D lags and variation across sectors. We show there is significant variation in research lags across sectors and a general trend towards lengthening R&D lags. This calls into question the assumption of a constant relationship between R&D expenditures and TFP over time and across sectors. Work in this paper also provides evidence that may be helpful in constructing new time-varying and sector-specific lag weights.

Third, we present evidence that agriculture typically relies on older ideas than the rest of the economy. In order to leverage the large body of work on the economics of innovation to the agricultural sector, it is important to understand the ways in which agriculture is distinctive and the ways in which it is not. In addition to summary statistics at the level of different agricultural subsectors, we also attempt to replicate the findings of Mukherjee et al. (2017), which links the value of patents and journal articles to the distribution of the age of cited references. Mukherjee et al. (2017) find a distribution centered on "young" ideas but with a high variance (e.g., a long tail of older ideas) is most associated with valuable ideas. Our results are generally consistent with this view, but when we estimate separate parameters for each subsector, we find some agricultural subsectors do not fit this pattern. In these cases, older ideas are frequently more valuable.

Finally, we present trends on the "expected" number of citations *received* versus the actual number of citations received, per year, between 1976 and 2016. This is a metric for the value of R&D conducted across different eras, and we generally find research conducted prior to the year 2000 is cited at a higher rate than we would expect. This trend exists for both patent and non-patent references and suggests post-2000 research may have been less useful at generating new ideas than pre-2000 research. This is potentially related to the debate over whether there has been a slowdown in agricultural TFP growth (Fuglie et al. 2017).

Our paper is organized as follows. We begin in section 1 with a description of our data sources and how we constructed agricultural datasets. Section 2 presents our evidence on the length of research lags in agriculture, including their variance across sectors and over time. Section 3 presents our econometric results linking the distribution of ages for patents to proxies for the value of patents. In section 4 we discuss how we construct "expected citations" received per year and compare expected to actual citations received. In section 5, we discuss our work in the light of some related literature. Section 6 concludes with some discussions of shortcomings in our research and potential policy applications.

1. Data

This paper is based on datasets covering 6 distinct agricultural subsectors: agricultural machinery, animal health products, biocides, biofuel products and processes, plant cultivars, and other ag. We begin by briefly describing how each dataset is constructed (full details are included in supplemental materials).

1.1. Defining Agricultural Subsectors

Three datasets were constructed by using the US patent and trademark office (USPTO)'s cooperative patent classification (CPC) system. These include the agricultural machinery, biocides, and "other ag" databases. The CPC system is designed to help patent examiners search for potentially antecedent technologies that a new application may be infringing upon, or is building upon. Patents are assigned one or several CPC codes to facilitate search for patents with similar technological elements. For these three datasets, our primary source is the current CPC codes files available from the USPTO's PatentsView website.

We designate agricultural machinery patents as those with a primary CPC code belonging to an agricultural machinery category and granted between 1976 and 2016. A full list is available in the supplemental materials, but major categories include subcategories A01B (soil working), A01C (planting, sowing, fertilizing), A01D (harvesting, mowing), A01F (processing of harvested produce), A01J (manufacture of dairy products) and A01M (adaptations and arrangements for liquid and powder-based spraying of biocides). Within these categories, we exclude animal-driven and hand-driven implements. The top assignees in this category are Deere & Company, CNH America, and Sperry Corporation, all of whom are major manufacturers of agricultural equipment and vehicles.

The biocide database is intended to encompass patents for herbicides, fungicides, insecticides, pesticides, and other chemicals used to control biological pests. It is not intended to include all chemical products in agriculture (such as fertilizers and animal drugs). As in the agricultural machinery categories, we include patents with a primary classification related to biocides and granted between 1976 and 2016. Biocides fall under CPC subcategory A01N 25/00 – A01N

65/48. The top assignees in this category are major chemical manufacturers such as Bayer, BASF, and Sumitomo chemical company.

The other agriculture category is a remainder category. It includes all patents with a primary CPC code falling under A01 (and excluding some categories related to fishing, hunting, and forestry), but which are not included in the other 5 datasets. Prominent types of technology in this category are hand implements for agriculture and horticulture, structures for the storing of crops and livestock, traditional plant breeding techniques, and a variety of devices for feedings, cleaning, managing, and caring for animals.

The biofuels dataset was developed by Albers et al. (2017), and their supplemental materials include extensive documentation. The goal of their project was to create a database of all patents for inventions filed around the world, over 1970-2013, that are involved in producing a liquid fuel from a renewable biological feedstock. They use a mix of keyword searching, international patent code classifications, specialized biofuel company names and manual examination to create their dataset. The authors generously shared this dataset with us. We use only the subset of patents granted in the USA over 1976-2013 in our dataset. The top assignees in this dataset include Pioneer Hi-Bred and Monsanto (patented plants that are major biofuel input stock are included in the dataset), but also Celanese International Corporation (an industrial chemical company) and Novozymes A/S (a producer of industrial enzymes, microorganisms and biopharmaceutical ingredients).

Plant cultivars have a variety of forms of available intellectual property rights protection. Some varieties of plant are eligible for protection from a specialized instrument called a plant patent, created in 1930. This instrument differs in several ways from utility patents (the most well known type of patent) and is not the focus of this paper. To avoid confusion, throughout the paper, we will refer to utility patents for plant cultivars as plant cultivar patents and cultivar patents, to avoid implying that we are talking about the specific instrument of plant patents.

Plant cultivars have only been eligible for utility patent protection since the 1980 Supreme Court decision Diamond v. Chakraborty. As such, our plant cultivars patent dataset extends only back to 1983, when the first such patent for a plant cultivar was granted (as far as we can ascertain). The plant cultivars database was assembled by Paul Heisey at USDA Economic Research Service, and was shared with us. In recent years, this dataset was assembled by searching for patents with CPC classification A01H, but excluding A01H 9/00 – A01H 17/00 (which correspond to non-agricultural organisms such as mosses, algae, and fungi). Earlier years used the US patent classification system (discontinued in 2015). Each result was then manually inspected to insure it corresponded to a patent for a plant cultivar (and not, for example, a patent for a plant breeding technique). The cultivar patent sector is unusually concentrated, with the two top companies (Monsanto and Pioneer) accounting for more than 60% of patents.

Finally, the animal health products dataset was assembled by us from archival FDA records. All veterinary drugs that make a medical claim must receive market approval from the FDA before they can be sold. As a consequence of the 1988 Generic Animal Drug and Patent Term Restoration Act, the FDA has published an annual list of all currently approved drugs (the “green books") since 1989. This list includes information on all patents associated with approved drugs (to help potential generic entrants identify unpatented drugs that could be good generic drugs candidates). We obtained every green book from 1989-2015 and extracted associated patents. Note that this includes the patents of drugs approved prior to 1989, so long as the drug approvals were still active in 1989.

This selection strategy means our animal health dataset has important differences from the other agricultural subsectors. In particular, we only observe patents associated with successful animal drug applications, not patents associated with all of animal health (including failed applications). This is unfortunate but inevitable. We know of no other reliable way for separating drug patents for human and animal health. Most of these patents are classified according to their chemical structure, not their use in veterinary medicine, and their titles and abstracts do not indicate whether their intended use is primarily veterinary. Indeed, many drugs are likely used in both markets. Moreover, most veterinary drugs are developed by a specialized animal health division in a larger human health company, but this is not reflected in the patent assignment. Indeed, the top animal health drug assignees include Pfizer, Eli Lilly and Company, Bayer, and Merck, all of which are primarily known for their work in human medicine.

Lastly, as a benchmark, we also frequently use the set of all US patents granted between 1976 and 2016. Figure 1 presents the annual number of patent grants by year in each subsector (log-scale), as well as the total number of patent grants by year. Over the 40 years we observe, plant cultivar patents and biofuel patents both experience substantial growth. Other sectors experience some growth, but do not change by an order of magnitude.

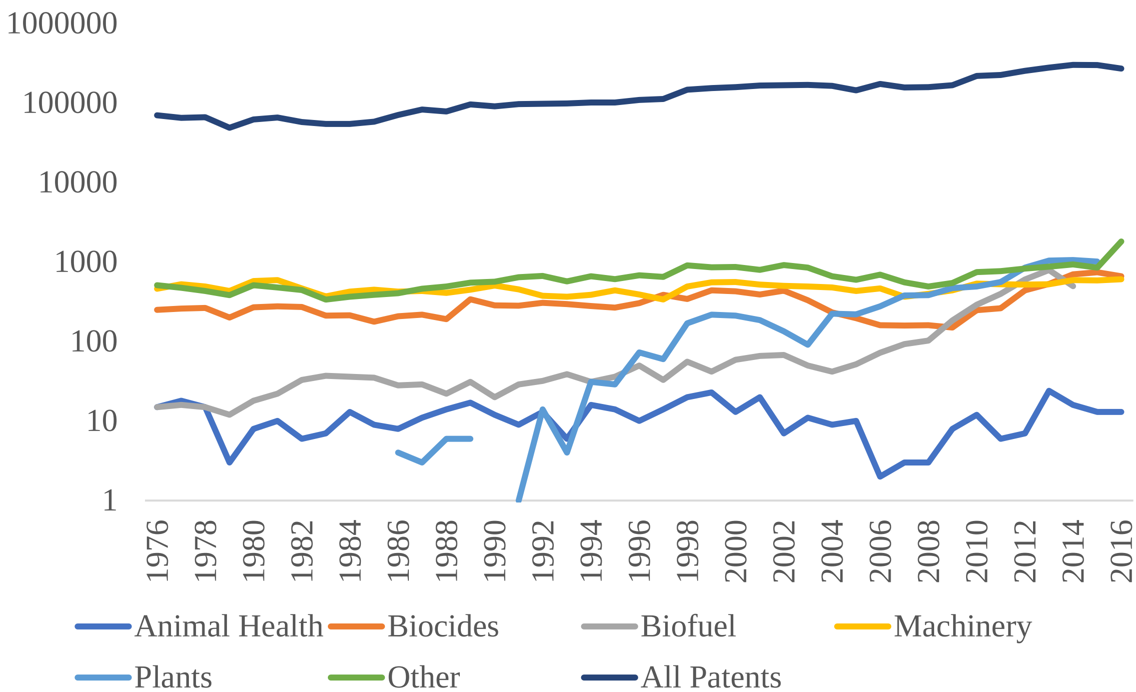

Figure 1. Annual patent grants by agricultural subsector (and all patents)

Table 1 provides a summary of how many patents are present in each dataset, and how many patents are cross-listed in multiple subsectors. The largest cross-listing is between biofuels and plant cultivar patents: 172 biofuel patents (4.2%) are plant cultivars. All other cross-listing are under 10 (and 1.8%). The most surprising listing are the two patents listed under both ag machines and plant cultivars. These appear to be plant cultivars that have been classified under CPC code A01C 11/00 ("Transplanting machines"). There are seventy patents with this classification, but the vast majority are planting machines and devices.

	Ag Machines	Animal Health	Biocides	Biofuels	Plants	Other Ag
Ag Machines	19,362	0	0	3	2	0
Animal Health	0	468	8	0	0	0
Biocides	0	8	12,774	7	0	0
Biofuels	3	0	7	4,087	172	0
Plants	2	0	0	172	8,206	0
Other Ag	0	0	0	0	0	26,844

Table 1. Cross-listed patents

1.2. Patent Characteristics

Once we have defined sets of patents belonging to different agricultural subsectors, we exploit two main categories of information: patent references and proxies for patent value. We will begin with a discussion of patent references.

1.2.1. References

For patent citations, we use the "uspatentcitation" file from the USPTO's PatentsView website. This provides a comprehensive list of all citations made by patents to other US patents. Patent grants are assigned numbers in the order they are granted, and so the year a patent is granted can be inferred from its number (see https://www.uspto/gov/web/offices/ac/ido/oeip/taf/issuyear.htm). Mean citations to patents by subsector are displayed in Table 3.

For non-patent references, we use the "otherreference" file from the USPTO's PatentsView website. Non-patent references are a far more heterogeneous group than patent citations. To get a better sense of what is included under the heading "non-patent reference" we performed an audit on 100 randomly selected references from each subsector, and from a sample of all patents. The results of this audit are displayed in Table 2.

	Ag Machines	Animal Health	Biocides	Biofuels	Plants	Other ag	All patents
Academic Journals	9	73	52	68	42	55	44
Conference Papers	7	4	6	0	3	3	5
Other tech. manuscript	3	1	4	1	3	3	3
Books	0	8	4	8	18	3	4
Trade magazines	10	1	4	0	0	2	2
Plant Breeder's Rights	0	0	0	0	17	2	1
Chemical abstracts	0	5	8	0	0	0	2
Patent related	33	2	13	11	12	12	18
Legal	0	0	0	1	0	3	1
Other	38	6	9	11	5	17	20

Table 2. Non-patent references audit

Academic journals represent the largest share of non-patent references, except in ag machinery. Weighted by the total number of references in each subsector (see Table 3), academic references account for an estimated 52% of all non-patent references in agricultural patents. Weighted by the total number of references, the next largest category is patent-related references (12.9% of references). This category primarily refers to patent application related documents and patent application search reports. Books account for an estimated 7.4% of ag references, plant breeders' rights for an estimated 4.2% of ag references, and conference papers for 3.2%. The "other technical manuscripts" category groups government and corporate reports, as well as unpublished manuscripts and working papers. It accounts for another estimated 2.8% of ag references.

We take this audit as generally supportive of our approach. Most of these non-patent references quite plausibly represent knowledge flows. Moreover, in most cases, the year of publication is a plausible proxy for the year relevant R&D was conducted. The combination of academic journal citations, conference papers, technical manuscripts and books (most of which are technical in nature) represent a clear case of knowledge flow from the research community and accounts for an estimated 65.7% of references. Citations to plant breeder's rights, chemical abstracts, and patent related documents seem at least as relevant as citations to other patents, and these account for another 19.0% of references.

The "other" category includes a mix of product descriptions, product manuals, websites, advertisements, articles in the popular press, and various uncategorizable products. Especially in the ag machinery category, product descriptions and manuals are a prominent reference type. These, and trade magazines, are less obviously proxies for knowledge that is used by the patent, but we do not discount the possibility that they are. It seems to us most likely that legal references do not represent true knowledge flows (and may instead be cited to support the patentability of an application), but these account for only an estimated 1.1% of citations. In future work, we hope to separate out the different forms of reference more closely. For this paper, we treat all non-patent references identically as useful proxies for R&D lags.

A second issue with non-patent references is determining the date of their publication. Non-patent references are not standardized in the way patent citations are (each patent citation includes the patent number, which links to all associated information). Instead, they are entered in a format of the patentee's choosing, which can vary even for the same reference. With 369,442 non-patent references, we did not have the resources to read each reference and find the appropriate year. Instead, to estimate the year associated with a non-patent reference, we scanned the text of each entry for a number between 1900 and 2017. If we find one such number, or multiple such numbers that all match, we assume this is the year the non-patent reference was published. So, for example, in the following reference, we would assign the year 1951, as this is the only numeral lying between 1900 and 2017.

> Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. Protein measurement with the folin phenol reagent. *J. Biol. Chem.* 193, 265-275 (1951).

However, if the page numbers had been 1900-1910, or if a slip of the fingers had rendered the volume number as 1932, then we would have multiple non-matching numerals in the range 1900-2017. If this is the case, we would have been unable to assign a year to the reference.

Using this methodology, we are able to assign a year to 87% of non-patent references (a low of 75% in ag machinery, and a high of 91% in plant cultivars). In 8% of references there is no numeral between 1900 and 2017, and in 5% of references there are multiple non-matching numbers between 1900 and 2017. In our audit, we verified our automated year extraction system correctly identifies years. The most common error is the failure to find a year when in fact the information is available. Table 3 provides some summary statistics on the number of non-patent references by subsector.

In this paper, we define the "age" of a citation as the gap between the grant date of the citer and cite. For example, if patent 999999 is granted in 2010 and cites patent 111111, which is granted in 1990, then the age of the citation is 20 years. Similarly, if a non-patent reference is published in 1990 and is cited by patent 9999999, then the reference's age is 20 years.

We close with an example. Consider US patent 8,544,574, titled "grain cart capable of self-propulsion" granted to Deere & Company in 2013. We would anticipate that this agricultural technology draws on a mix of new and old ideas. Self-propulsion, as part of the self-driving technology program, is a recent innovation. However, the idea of a grain cart is quite old, and we should not be surprised if the optimal design was fixed a long time ago.

The patent has 32 datable references (and 1 non-patent reference missing a date), 22 of which are to patents, 9 to patent-related documents (such as applications), 1 to a conference proceeding. Of these, fully a third (11) are dated in 2010 or later, indicating the product draws heavily on very recently developed technologies and research. Another 8 references (including a citation to the proceedings of a 2007 IEEE symposium on intelligent vehicles) range between 2000 and 2009. This is intuitive; self-driving technology is a relatively recent innovation, and so it is not surprising that over half of the references are from 2000 and after. At the same time, the grain cart, as a vehicle, is a very old idea, and this is also reflected in the patent's long tail of older references. Another 10 citations are from 1950-2000, and the patent makes 5 references to patents older than 1950 (the oldest is 1928).

1.2.2. Proxies for Patent Value

Patents vary substantially in their value. Most are worth little, but a few are very valuable. Understanding the factors that drive the most valuable patents, which may be much smaller in number, is therefore important. However, because we generally do not observe market prices for patents, we must use proxies for their value. We use three different proxies of value: renewal data, Kogan et al. (2017) estimates of patent value, and citations received within 5 years.

Renewal data provides our first proxy of value. Since 1980, US patent-holders have had to pay escalating fees to keep their patents in force after 4, 8, and 12 years. There is a large literature that uses patent renewal decisions as a proxy for the value of patents (Pakes and Schankerman 1984, Lanjouw et al. 1998, Baudry and Dumont 2006, Bessen 2008, Serrano 2010, Clancy 2017). This literature assumes a rational patent-holder will only pay for renewal if the value of continued patent protection exceeds the fee. To maximally differentiate high-value patents, we use "fully renewed" as our indicator of value, which indicates the patent-holder paid the 12-year patent fee and renewed the patent through to the end of its life. Note that we can only compute this for patents granted after 1981 (when fees came into effect) and prior to 2005 (because we need 12 years to observe renewal fees).

Our second proxy of value is Kogan et al. (2017)'s estimates of the value of patents held by publicly traded firms. We denote this measure KPSS after the surnames of the authors. Kogan et al. (2017) base their estimates on the movements in market capitalization of the patent-holder in the three days before-and-after the patent's grant is announced. These observations are available throughout the entire period, but only for the (typically small) sample of patents that belong to publicly traded firms at the time of their grant.

Our final proxy of value is the number of citations received by a patent in the subsequent five years. There is a large literature that uses citations received as a proxy for patent value (Nagaoka, Motohashi, and Goto 2010). To give each patent 5 years in which to accumulate citations, we only have citation estimates for patents granted prior to 2012.

These three proxies vary in their interpretation. Renewal data and KPSS both correspond relatively tightly to different evaluations of the private value of a patent. The renewal indicator is very coarse (patents are either renewed or not), but reflect the patent-holder's subjective judgement about the value of continued patent protection after 12 years of data. KPSS value is continuous and represents the collective judgment of market participants. However, this decision is made at the time of a patent's grant and must be extracted against the background of noisy market fluctuations. Moreover, when multiple patents are granted on the same day to the same firm, Kogan et al. (2017) equally apportion the total change in the publicly traded firm's value among the patents.

Patent citations, in contrast, arguably reflect the social value of an invention as well as the private value of a patent. Indeed, this paper is premised on the idea that patent citations reflect knowledge spillovers. Highly cited patents are ones that generated ideas useful to many other patents. This value is not necessarily captured by the patent-holder though. At the same time, citations can also indicate the patent occupies valuable real estate in technological space, prompting other firms to enter the same space (and drawing on the ideas contained in the originator). This is a proxy that patents in this domain have high private value (or else the private sector would not enter).

1.3. Data Summary

Table 3 presents summary data on the 6 agricultural subsectors we consider.

	Ag Machine	Animal Health	Biocides	Biofuel	Plants	Other Ag
Patents	19,362	468	12,774	4,087	8,206	26,844
Start Year	1976	1976	1976	1976	1983	1976
End Year	2016	2016	2016	2014	2015	2016
Citations made (mean)	13.1	8.2	8.3	16.5	6.9	11.9
Non-patent references (mean)	0.8	6.5	5.6	17.0	8.4	3.5
Fully renewed	38.6%	84.6%	40.7%	36.0%	88.2%	26.2%
KPSS (mean)	8.5	34.5	8.6	15.5	19.6	12.2
Citations received (mean)	1.6	2.3	1.0	3.0	2.5	1.4

Table 3. Summary Statistics

Before we make a few observations, it will be useful to introduce two major categories of technology: hardware-based and biochemical based. Broadly speaking, the ag machinery and other ag categories relate primarily to hardware such as physical machines, devices, and implements. In contrast, the plant, animal health, and biocide categories relate primarily to biological and chemical processes. The biofuels category is a mix of both hardware (in the form of furnaces, fermenters, boilers, and other hardware for the production of biofuels), and biochemical products (such as plant cultivars, enzymes, and other chemical inputs). Throughout the remainder of this paper, we will note that hardware and biochemical subsectors frequently exhibit notable differences.

First, with the exception of the biofuel sector, most patents make 14-16 total references (patent and non-patent), but the share of these varies substantially. Among the biochemical subsectors, more than 40% of references are not to patents. For the hardware subsectors, less than 25% are not to patents. Biofuel patents make more patent citations and more non-patent references than any other sector, and have more than 50% of citations to non-patent references.

Second, by the fully renewed share and KPSS value proxies, animal health patents and plant cultivar patents are most valuable. This is not so surprising for animal health for two reasons. First, as discussed in section 1.1, this is a highly selected sample consisting only of patents associated with drugs that received market approval. Second, the pharmaceutical sector is typically one in which patents are rated as very important (Arora, Ceccagnoli, Cohen 2008). The high value of cultivar patents is more surprising and warrants further study.

2. Research Lags

The first question we address is the distribution of research lags, both in agricultural as a whole, and across subsectors and time. To begin, we present the probability mass functions and cumulative mass functions for the age of cited references (pooling patent citations and non-patent citations). In each figure, we present results for ag patents and for all patents for ages 0-50 years.

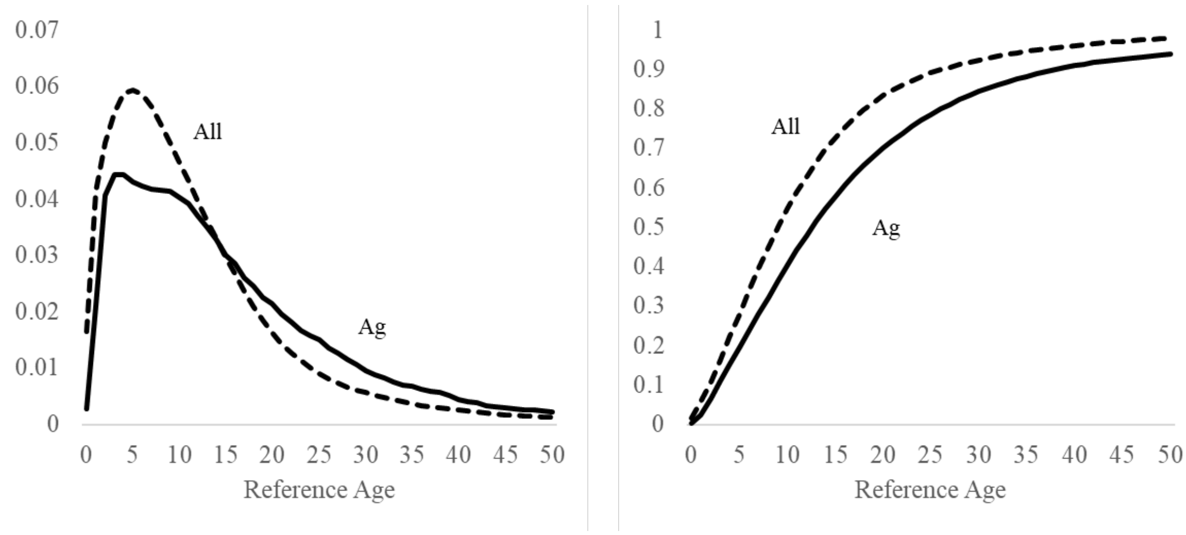

Figure 2. Distribution of reference ages by year (left) and cumulative (right)

Note: Vertical axis is the share of references made to documents of a given age

Clearly, taken as a whole, agriculture relies on older ideas than they typical patent. This result holds when considering citations to patents and non-patent references in isolation (see appendices), though the difference is stronger for patents. Table 4 presents a large set of disaggregated summary statistics.

	Mean			Median			Mode			St. Dev		
	NPR –	All –	Pat	NPR –	All –	Pat	NPR –	All –	Pat	NPR –	All –	Pat
Plants	14.7	**12.5**	9.8	13	**11**	9	2	**2**	3	10.8	**9.5**	6.6
Animal Health	15.0	**13.8**	12.9	12	**11**	10	13	**10**	8	11.1	**10.5**	9.9
All	11.3	**14.1**	15.0	9	**10**	11	5	**5**	5	9.2	**13.9**	15.0
Biocides	14.3	**15.0**	15.5	11	**12**	13	6	**7**	7	12.0	**12.1**	12.2
Biofuel	13.3	**15.7**	18.1	11	**12**	14	8	**8**	10	10.5	**13.1**	15.0
Ag	13.4	**18.6**	20.7	12	**13**	14	2	**4**	4	10.7	**18.9**	21.0
Ag Machines	8.1	**21.4**	22.2	5	**15**	16	4	**4**	5	8.5	**21.2**	21.5
Other	12.6	**21.1**	23.6	11	**14**	15	10	**5**	5	9.5	**22.4**	24.4

Table 4. Summary statistics on reference age regressions

Notes: Presents summary statistics on the distribution of reference ages for all subsectors, broken down by non-patent references (NPR), patents (Pat), and pooled (All).

We present four summary statistics: mean, median, mode, and standard deviation. Within each of these, we present the statistic for non-patent references ages (the NPR column), patent references (the Pat column) and pooled references (the All column). Statistics for the pooled reference ages are in bold. Rows have been sorted by the mean reference age for pooled references (the bolded mean column).

We begin by restricting our attention to only the bolded numbers (which correspond to references to patents and other documents). Even though ag patents rely on older ideas *on average*, when we disaggregate, we can see this is not true of every subsector. Cultivar patents and animal health patents both have lower mean ages and cultivar patents also have a very low mode. Note also that each sector has a lower standard deviation then all patents, suggesting a shorter right tail.

Every subsector, however, has a higher median reference age than all patents. Moreover, the biocides, biofuel, ag machines, and other ag sectors all have higher means, higher medians. Most also have higher standard deviations, but both ag machines and other ag have modes as short or shorter than all patents. This suggests the high mean and median age of their references is largely about the length of the right tail.

Without disaggregating the references into patent and non-patent categories, we might conclude that most ag subsectors draw on older ideas than typical patents, with animal health and plants an exception. However, when we divide up the references into patent and non-patent references, we see that the plant and animal health categories are unusual in another way: their mean, median, and standard deviation of non-patent references are higher than for patent references. This is in contrast to every other sector.

In the case of cultivar patents, there are strong reasons to believe this is an artifact of changes to patent law. As noted in section 1.1, prior to 1980, many forms of biological innovation were not eligible for patent protection. This changed in 1980, with the Supreme Court decision *Diamond v. Chakraborty*, but the new rules for what constituted patentable subject matter continued to be clarified through the following decades (including the 2001 decision *J.E.M. Ag Supply, Inc. v. Pioneer Hi-Bred*). Concretely, this means, for example, there were no plant cultivar patents to cite prior to at least 1980. Plant cultivars may well have drawn on plant breeding initiated before 1980, but this contribution would not be reflected in patent citations. This can explain why the mean age of cultivar patents is so low: whereas other sectors can cite patents as early as 1836, cultivar patents can only cite patents as old as 1980. Since this restriction doesn't apply to non-patent references, it's non-patent references are higher. A similar artifact may be occurring in animal health, if it relies on the kinds of biological patents that were only eligible for patent protection after 1980.

Note that the plant and animal health subsectors both have non-patent reference age mean, median, and standard deviations *above* the typical patent. Because these are less likely to be impacted by truncation issues, these results seem more representative of the age of ideas used in these subsectors. This suggests then that every agricultural subsector uses older than typical patents, on average.

> **Stylized Fact 1a:** Over 1976-2016, agricultural patents in all sectors tended to cite older patent and other documents than non-agricultural patents.
>
> **Stylized Fact 1b:** The tendency of agriculture to cite older ideas is driven primarily by the persistence of old ideas, rather than delays in citing young ideas.

A second observation that emerges from Table 2 is the heterogeneity across agricultural subsectors.

The mean patent citation age for hardware categories is significantly longer than for biochemical sectors, even when we restrict attention to the biocide dataset (which is less likely to be impacted by the 1980 change in patentable subject matter). Furthermore, in hardware subsectors, there is a larger divergence in the mean age of patent and non-patent references.

> **Stylized Fact 2:** Hardware-based agricultural subsectors tend to cite patents significantly older than their non-patent references and also older than the patents cited by the biochemical-based agricultural subsectors.

3. Trends over time

We next turn to the question of whether research lags are changing over time. Figure 3 illustrates the mean age of citations made by patents granted in different years, across different sectors. We present the centered 5-year moving average of mean age to smooth out year-to-year noise and focus on larger trends.

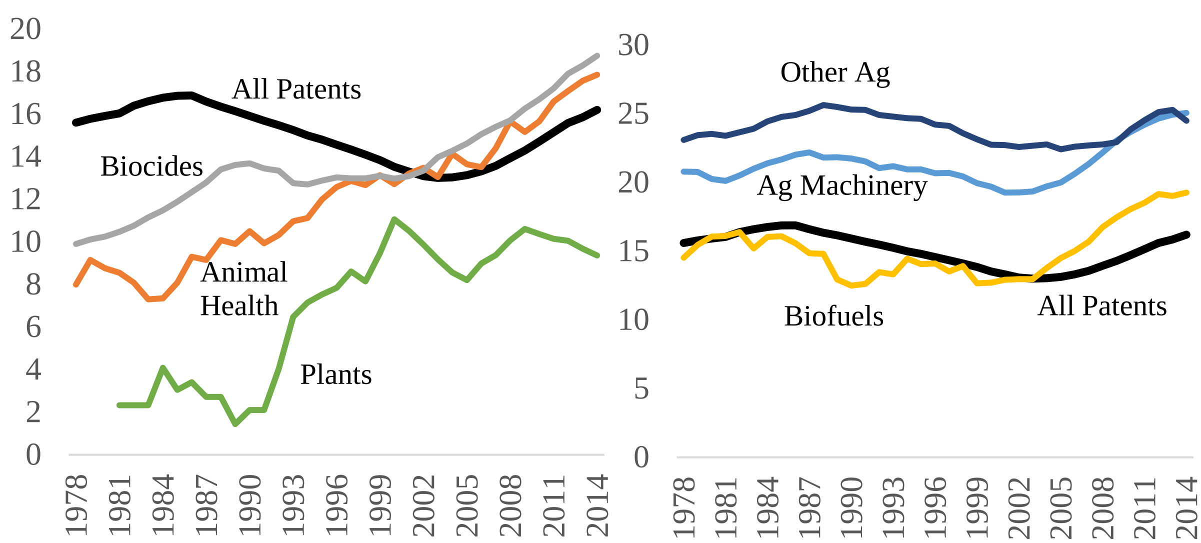

Figure 3. Mean Age of Patent Citations (vertical axis) by Citing Patent Grant Date (horizontal axis): 5-year moving average

The first thing to note is a trend towards the use of older ideas in later years across all sectors. This is especially pronounced for biochemical ag subsectors (Figure 2 left), but is also present in the hardware and biofuels subsectors (Figure 2 right), to some degree, in the 2000s. Notably, by the end of the period, every ag sector except for plant cultivar patents has a higher mean age than the typical patent, further bolstering stylized fact 1. Leaving aside plant cultivars, the typical age of patent citations ranges from 17-25 years.

Figure 4 presents the mean age of non-patent references, by the grant year of the citing patent. Once again, we use 5-year moving averages to highlight trends instead of year-to-year variation.

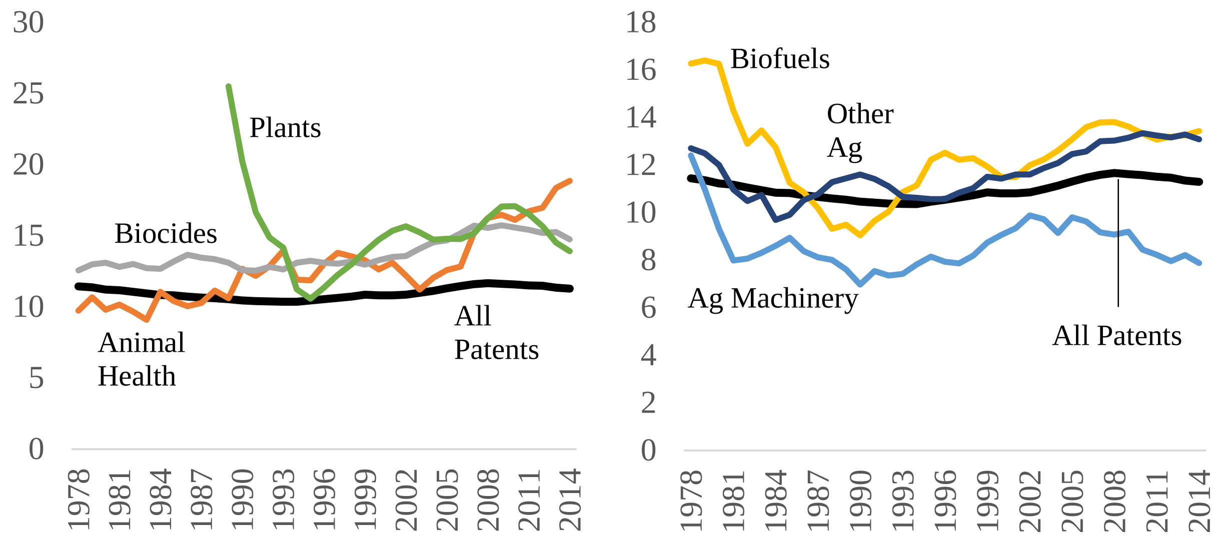

Figure 3. Mean Age of Non-Patent References (vertical axis) by Citing Patent Grant Date (horizontal axis): 5-year moving average

Several trends are again notable. Once again, we observe a general trend towards older reference ages among biochemical sectors (Figure 3 left). Cultivar patents appear to be an exception, but the steep decline in mean non-patent reference age for plants through the early 1990s is driven by a very small number of patents (there were only 15 granted prior to 1990). Among hardware subsectors, we see the pattern of initially decreasing age, followed by increasing age that also characterized the mean age of patent citations (although ag machinery has a decreasing mean reference age toward the end of our sample).

> **Stylized Fact #3:** The mean age of the ideas used in agricultural patents has been increasing for at least a decade in most subsectors. The age of ideas used in biochemical subsectors have been increasing for several decades.

4. Patent Value

While the preceding stylized facts are useful, because of the strong heterogeneity of patents, it is desirable to see if the vintage of ideas used is correlated with proxies for the value of patents. In other words, even though agriculture tends to use older knowledge than typical patents, what really matters is what the best R&D in agriculture uses. To explore this issue, we regress proxies for patent values on statistics on the distribution of age in a patent.

In particular, we follow Mukherjee et al. (2017) in constructing variables about the vintage of ideas: the mean age of a patent's citations and non-patent references, and the coefficient of variation of a patent's citations and non-patent references. Mukherjee et al (2017) find a "sweet spot" in the vintage of knowledge cited by papers, patents, and other documents. Time after time, valuable ideas have a low mean age but a high coefficient of variation. Thus, creative work primarily drawing on recent advances but also including a long tail of older ideas tends to be most valuable by a number of proxies.

We separate out the mean and coefficient of variation for patent citations and non-patent references because they appear to measure different things. As Table 5 indicates, patent variables and non-patent variables tend to be weakly negatively correlated with each other:

	Mean Age – Patent Citations	Mean Age – Non-Patent References	Coefficient of Variation – Patent Citations
Mean Age – Non-Patent References	-0.17		
Coefficient of Variation – Patent Citations	0.24	-0.18	
Coefficient of Variation – Non-Patent References	-0.18	0.52	-0.13

Table 5. Age Correlations

For each proxy for value, we run a regression with dummy variables for each year and for each subsector. Because the coefficient of variation of the distribution of patent ages may be correlated with the number of citations a patent makes, and because the latter has been shown to be correlated with patent value, we also include the inverse-hyberbolic sine of citations made. Lastly, we include a dummy variable for the presence of a KPSS estimate, as this is an indicator that the patent belonged to a publicly traded firm when granted. All of these controls tend to be statistically significant.

The functional form of our regressions varies across the proxy for value. Renewal is a binary variable equal to 1 if the patent paid its 12-year renewal fee and 0 otherwise. We use a logit regression to predict the probability a patent is fully renewed. Our KPSS variables are a simple OLS regression over the log values of the KPSS estimate of value. Finally, because citations received is an integer truncated at 0, we use a quasipoisson regression to link citations received with our explanatory variables.

Our pooled results are for all agricultural patents are presented in Table 6.

Dependent Variable	**Renewal**	**KPSS**	**Citations**
Intercept	0.48*	1.47***	0.04
	(0.23)	(0.14)	(0.09)
Biocide	-2.13***	-1.11***	-0.78***
	(0.18)	(0.11)	(0.07)
Biofuel	-2.27***	-0.94***	0.17*
	(0.19)	(0.14)	(0.08)
Ag Machines	-1.99***	-1.05***	-0.37***
	(0.18)	(0.11)	(0.07)
Other Ag	-2.54***	-1.19***	-0.44***
	(0.18)	(0.12)	(0.07)
Plants	0.08	-0.36**	0.02
	(0.19)	(0.12)	(0.07)
IHS(cites made)	0.11***	0.14***	0.28***
	(0.01)	(0.02)	(0.01)
Publicly Traded	0.62***		0.21***
	(0.03)		(0.02)
Mean – patent citation	-0.01***	0.00	-0.02***
	(0.00)	(0.00)	(0.00)
Mean – non-patent ref	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)
C.V. – patent citation	-0.04	-0.04	0.06***
	(0.03)	(0.04)	(0.02)
C.V. – non-patent ref	0.56***	0.08	0.13***
	(0.07)	(0.05)	(0.01)
Num. obs.	34,393	10,013	54,533

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors in parentheses. All regressions include year fixed effects

Renewal: logit regression over probability patent is renewed to term (1982-2005)

KPSS: OLS regression over log value of Kogan et al. (2017) estimates of patent value (1976-2016)

Citations: Quasipoisson regression over number of citations received in 5 years after grant (1976-2012)

Table 6. Patent Value and Idea Vintage (all patents)

The results depicted in Table 6 broadly support the claims of Mukherjee et al. (2017). We find the coefficient on “mean patent citation age” is negative and statistically distinguishable from zero when we use renewal and citations received as proxies for value. The coefficient on “coefficient of variation – non-patent reference” is positive and statistically distinguishable from zero in both cases. Moreover, when we use citations received (the proxy emphasized in Mukherjee et al. 2017), the estimated coefficient for “coefficient of variation – patent citation” is also positive and statistically significant. We do not detect any statistically significant results for KPSS proxies, but these are also the ones for which we have the fewest observations. Taken together, the results are weakly supportive of the idea that valuable patent in agriculture tend to have a distribution of idea ages centered on younger ideas, but with a long tail of older ideas also used. In the appendix, we perform robustness checks using alternative functional forms and obtain broadly similar results.

Table 7 replicates this exercise for each agricultural subsector, so that we can allow for subsector-specific parameters. This is a dense table that summarizes four coefficients of interest across 18 separate regressions. Columns display regression results for a given sector. Rows summarize results for a given dependent variable. Finally, like coefficients are grouped together in four sub-tables. For example, the estimated coefficient on the mean patent citation age, for animal health, when using citations received as our proxy for value, is given in the third row of the first column, under the sub-table heading "Mean – Patent Citation Age." Bolded coefficients are statistically significant at the 5% level.

If our results were maximally consistent with Mukherjee et al (2017), all coefficients under the headings "Mean – Patent Citation Age" and "Mean – Non-Patent Reference Age" would be negative, while all coefficients under "Coefficient of Variation – Patent Citation Age" and "Coefficient of Variation – Non-Patent Reference Age" would be positive. This would indicate using younger ideas on average, but with a long tail of older ideas, is correlated with higher patent value, regardless of whether ideas are proxied by patent citations or non-patent references.

Mukherjee et al. 2017 emphasize receipt of citations as their measure of value. Setting aside patents for plant cultivars, when value is measured by citations received, our results are consistent with their findings. In every one of the non-plant sectors, we find evidence (not always statistically significant) that using younger ideas on average, but with a larger variance, is associated with patent citations.

This results begins to look a bit scrambled, however, when we look at alternative proxies of patent value. In the hardware subsectors (specifically, ag machinery and other ag), the probability of renewal to term, or KPSS estimated patent values, may be *positively* related to the average age of ideas used. Turning to the coefficient of variation, a narrower spread of ideas (which would indicate the absence of a long tail of older ideas) is sometimes associated with higher probability of renewal (other ag), or higher KPSS estimates of value (biofuels).

Lastly, for cultivar patents, we consistently estimate results that are somewhat contrary to Mukherjee et al. (2017). At least with cultivar patents, it appears the citation of older ideas, on average, is associated with more valuable patents. We find some contradictory results for the coefficient of variation in cultivar patents, depending on the measure of patent value and whether we look at patent citations or non-patent references.

Dependent Variable	Animal Health	Biocides	Biofuels	Ag. Machinery	Plants	Other Ag
			Mean – Patent Citation Age			
Renewal	-0.01 (0.04)	-0.00 (0.00)	-0.00 (0.01)	**-0.02** (0.00)	**0.13** (0.04)	**-0.01** (0.00)
KPSS	0.02 (0.02)	0.01 (0.00)	-0.01 (0.01)	**0.01** (0.00)	**0.03** (0.01)	**-0.01** (0.01)
Citations	**-0.07** (0.02)	**-0.01** (0.00)	-0.01 (0.01)	**-0.02** (0.00)	**0.01** (0.00)	**-0.02** (0.00)
			Mean – Non-Patent Reference Age			
Renewal	-0.04 (0.03)	**-0.01** (0.00)	-0.01 (0.01)	0.00 (0.00)	-0.02 (0.01)	**0.01** (0.00)
KPSS	0.00 (0.02)	**-0.01** (0.00)	-0.00 (0.02)	**0.02** (0.01)	**0.02** (0.00)	**0.02** (0.01)
Citations	0.01 (0.01)	-0.00 (0.00)	-0.01 (0.01)	0.00 (0.00)	0.00 (0.01)	**-0.01** (0.00)
			Coefficient of Variation – Patent Citation Age			
Renewal	-1.03 (0.83)	0.06 (0.06)	-0.42 (0.29)	0.13 (0.08)	**1.47** (0.64)	**-0.23** (0.07)
KPSS	-0.22 (0.46)	-0.00 (0.05)	0.58 (0.41)	0.14 (0.09)	-0.20 (0.18)	-0.32 (0.18)
Citations	0.27 (0.37)	0.03 (0.04)	0.31 (0.17)	0.05 (0.04)	**0.44** (0.21)	**0.14** (0.04)
			Coefficient of Variation – Non-Patent Reference Age			
Renewal	0.21 (0.87)	0.18 (0.10)	**0.93** (0.32)	**0.85** (0.20)	-0.02 (0.38)	**0.71** (0.13)
KPSS	-0.54 (0.57)	0.03 (0.06)	**-1.09** (0.42)	0.04 (0.22)	**0.51** (0.16)	0.36 (0.21)
Citations	0.44 (0.31)	**0.09** (0.02)	**0.87** (0.13)	**0.48** (0.07)	**-0.38** (0.17)	**0.13** (0.04)
			# of Observations			
Renewal	285	6768	909	10235	1659	14537
KPSS	201	3033	308	4051	1313	1107
Citations	392	9691	2187	16508	4220	21535

Notes: Bold coefficients indicates $p<0.05$. Standard errors in parentheses
All regressions include year fixed effects, inverse-hyberbolic sine of number of citations made, and a dummy if Kogan values are available.
Renewal: logit regression over probability patent is renewed to term (1982-2005)
KPSS: OLS regression over log value of Kogan et al. (2017) estimates of patent value (1976-2016)
Citations: Quasipoisson regression over number of citations received in 5 years after grant (1976-2012)

Table 7. Patent Value and Idea Vintage (subsector specific)

In the appendix, we present the results for alternative specifications, but this does not change our results as presented. We summarize this bundle of findings in three more stylized facts:

> **Stylized Fact 4:** For non-cultivar agricultural patents, a distribution of referenced idea ages with a younger mean and longer right tail is weakly associated with receipt of more citations.

> **Stylized Fact 5:** For non-cultivar agricultural patents, measures of patent value based on renewal or Kogan et al. (2017)'s estimates do not reveal a clear relationship between the distribution of referenced idea ages and patent value.

> **Stylized Fact 6:** For cultivar patents, an older mean referenced idea age is associated with more valuable patents.

5. Which era's ideas were most heavily used?

So far we have presented evidence on the age of the ideas used by patents, but have not speculated as to why there is variation in age across sectors and over time. In this section, we present an alternative framing based on the extent to which different eras' work is cited. It may be that the research conducted in different eras is more or less valuable as a foundation on which to build future research, and this may drive changes in the mean age of citations.

For example, foundational work in genetic engineering from the 1970s and 1980s might be cited for many subsequent decades, as applications are spun out. Alternatively, the digital revolution might form a platform for future work in the 1990s, to which hardware researchers return again and again. If this is the case, the mean age of ideas cited would lengthen, but only because firms return to the same foundational research in year after year, and these years get farther and farther in the past.

Defining how "heavily used" is the research of an era is conceptually tricky. A seemingly simple metric might be the total number of citations received by ideas from a given year. This is depicted in Figure 4, with the maximum number of references (of all types) received in a single year normalized to 100 for comparison across subsectors.

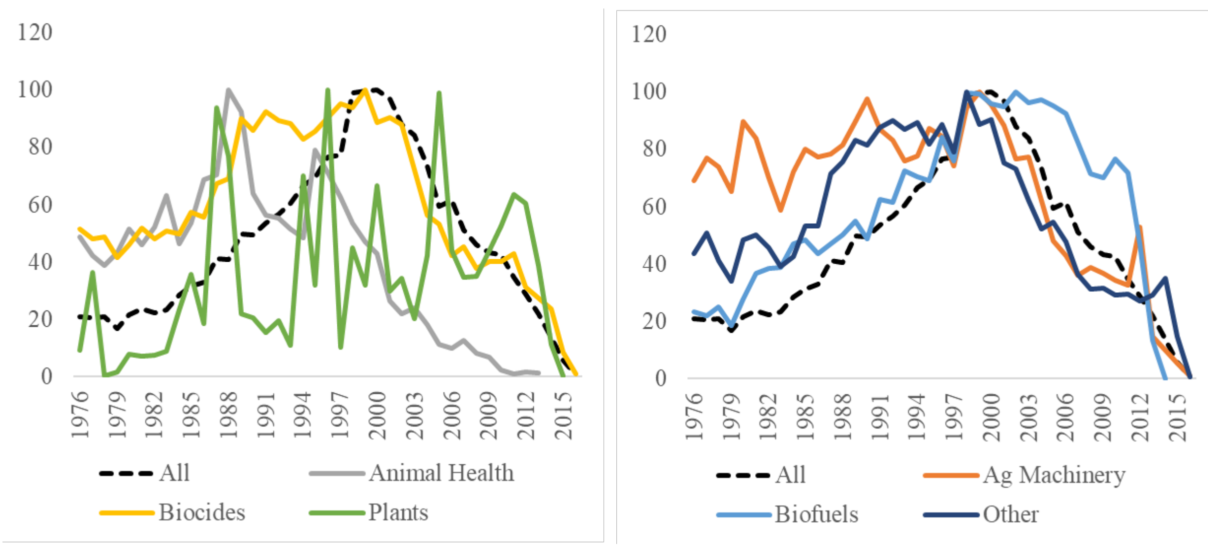

Figure 4. Citations received by year research is published (max citations received normalized to 100).

In this jumble of lines, we detect in most sectors an upward trend through the year 2000, followed by a downward trend. Cultivar patents represent an outlier, with several individual years receiving far above the normal number of citations. These correspond to years in which various highly cited articles and books were published. This clumping may be unique to cultivar patents because the majority of cultivar patents belong to just two companies: Monsanto and Pioneer Hi-bred. If these companies use a standard set of references in most of their patents, it could account for the high degree of clumping by year.

Turning our attention to the other subsectors and to patents as a whole, it is difficult to draw any conclusions because of a host of confounding factors. First, as we saw in Figure 2, most patents and other documents are only cited a few years after they are published, and then there is a long tail of citations to older documents. This will privilege older ideas over new ones, and probably accounts for the declining number of citations received by ideas after 2000. These ideas simply haven't been around long enough to accrue many citations.

On the other hand, Figure 4 is also complicated by two secular trends in patenting, neither of which is obviously driven by changes in the quality of research. First, the number of patents granted per year is growing, which implies newer patents will receive more citations than older ones over time. Second, the number of citations per patent is also growing. This will further augment the number of citations received by patents in the most recent year. These two trends likely account for the increasing number of citations per year received from 1976 to 2000.

To better highlight eras that are highly cited, we need to compare the number of citations received to the number that would be *expected*, given these trends. To compute the expected number of citations received by a year, we use our sector-specific distributions of reference ages for patents and non-patent references. Specifically, define:

> $s_{pi}(a)$: the share of a subsector i patent's citations to other patents with age a.
>
> $s_{ri}(a)$: the share of a subsector i patent's non-patent reference to references with age a.

Thus, if 10% of ag machinery's citations to patents go to patents with an age of 3, then $s_{pi}(3) = 0.1$ for i = ag machinery. Further define n_{ti} to be the number of patents granted in year t in subsector i, p_{ti} to be the average number of citations to other patents in subsector i in year t, and r_{ti} to be the average number of non-patent references in subsector i in year t.

Let y_{ti} denote the actual number of references received by documents published in year t by subsector i. The *expected* number of references received by documents published in year y is defined as:

$$\hat{y}_{ti} = \sum_{t=t_0}^{2016} n_{ti}\left(p_{ti} \cdot s_{pi}\left(t-t_0\right) + r_{ti} \cdot s_{ri}\left(t-t_0\right)\right) \tag{1.1}$$

Equation (1.1) controls for changes in the number of patents granted in each year with n_{ti}, changes in the number of references made by patents over time with p_{ti} and r_{ti}, and the bias towards older documents with $s_{pi}(a)$ and $s_{ri}(a)$. When y_{ti} and $\hat{y}_{ti}$ do not converge, this is because the distribution of reference ages is changing over time (note that $s_{pi}(a)$ and $s_{ri}(a)$ do not have a time subscript, but are estimated for all patents in a subsector, across all years). Thus, this section is essentially an alternative way of presenting the results in section 3, but emphasizing variation in which years' research gets cited.

Figure 5 presents the ratio of actual relative to expected citations, with the same figure for "all patents" also computed as a benchmark. When the line is above 1, this means documents published in the year received more than the expected number of references from subsequent patents. When the line is above the “all patents” line, it indicates documents published in the year had a higher actual/relative ratio than for patents as a whole.

In general, figure 5 suggests agricultural patents in the first half of the 1976-2016 period received more than the expected number of citations and fewer than the expected number late in the period. Both trends are common to all non-plant subsectors. “All patents” exhibit the same trends (above expected citation early, below expected citation later), but agricultural patents experience this divergence to a greater degree.

> **Stylized Fact 7a:** Documents published in the first (second) half of 1976-2016 are more (less) highly cited by non-cultivar agricultural patents relative to expectation.

> **Stylized Fact 7b:** While all patents exhibit the trend highlighted in fact 7a, ahe extent to which actual citations vary from expectation is greater for agricultural patents.

In the appendix we present complementary figures for just non-patent references and just patent citations. These trends are present in either case.

Figure 5. Actual relative to expected citations received by year research is published

Note: In all figures, the solid line corresponds to the relevant agricultural subsector and the dashed line to all patents.

6. Discussion: relationship with other literature

In this section, we offer some brief comments on some of the implications of these results for questions in the literature.

6.1. Research Lags

Table 1, adapted from Fuglie et al. (2017), presents a brief summary of common R&D lag assumptions. To compute the summary statistics mean lag and standard deviation of lag, we use the weights attached to each lag, normalizing the sum of weights to 1. In this literature, the mean lag of R&D in the construction of a knowledge stock ranges from 13.9 to 26.7 years, with standard deviations as narrow as 3.6 years and as wide as 8.3 years.

R&D Lag Structure	Papers	Mean Lag	St. Dev. of Lag
50-year gamma	Anderson and Song (2013)	26.7	8.3
	Alston et al. (2010)		
35-year trapezoidal	Jin and Huffman (2016)	15.2	7.3
	Wang et al. (2012)		
	Huffman and Evenson (2006)		
35-year inverted V	Wang et al. (2013)	22.6	3.6
19-year inverted V	Wang et al. (2013)	13.9	4.2
30-year non-parametric (public)	Chavos and Cox (1992)	22.6	3.6
30-year non-parametric (private)	Chavos and Cox (1992)	13.9	4.2

Table 8. Common Lag Structures

The mean age of citations is closely related to the weight put on R&D, as it signifies the length of time between when a (completed) research project bears fruit in the form of an agricultural input with some commercial value. Across agriculture as a whole, and combining all forms of reference, we find a mean lag of 18.6 years, which is squarely in the middle of the assumed ranges in Table 8. As we argued at the outset, this is more likely to be a lower bound on the typical lag between R&D expenditures and an impact on TFP for two reasons. First, technologies may take time to be adopted after they are patented, and only once they are adopted can they impact TFP (although it is also possible for technologies to diffuse while in a "patent pending" status). Second, R&D takes time to perform, and our estimates are based on time from when an R&D project has a published output.

A second issue is that most patents represent private R&D, and most of the R&D lags in Table 8 are estimated with respect to public sector R&D. One pathway for public sector R&D to impact TFP is via private sector adaptation and commercialization of public sector research. However, it is also possible for the public sector to directly develop agricultural inputs, such as increased plant breeds (note however, that plant breeding is increasingly a private sector activity – see Clancy and Moschini 2017). If this is the case, then the age of citations to other patents may not be a good source of comparison to the R&D distributions summarized in Table 8 (since it corresponds mostly to private adaptation of private sector R&D).

However, as indicated in Table 1, a large share of non-patent references are made to academic work. The public sector still funds the majority of academic work, and so our estimates of the mean age of non-patent references may be interpreted as more closely corresponding to private sector adaptation of public R&D. For agriculture as a whole, we estimated the mean age of non-patent references at 13.4 which is slightly below the minimum mean lag (13.9 years) given in Table 8. Once again, we note this is more likely a lower bound on the "true" lag.

While our estimates of the mean age of ideas used in agriculture is within the range of assumptions made about R&D lags given in Table 8, our results suggest these estimates under-estimate the extent to which old ideas continue to have life in agriculture. Looking only at non-patent references, we obtain a standard deviation of 10.7 years, compared to a max of 8.3 years in Table 8. When we turn to "all references" we obtain a standard deviation of 18.9 years. Moreover, as our results in section 4 illustrate, the importance of a long tail of older ideas is correlated with the most valuable patents (and not simply the number of patents).

Second, our results also demonstrate that R&D lags differ across types of innovation (particularly between biochemical and hardware based ag innovation), and also change over time. In particular, we find that the R&D lag has been getting longer in all fields for at least a decade, and has been getting longer for several decades in biochemical ag sectors. Between 1994 and 2014, the patent-weighted mean age of non-patent references for all ag increased by 2.6 years.

6.2. Productivity Slowdown

A long-running disagreement in the economics of agricultural R&D is the extent to which there is or is not a slowdown in the growth rate of agricultural productivity. Fuglie et al. (2017) present a recent overview of the disagreement. In this literature, it is sometimes suggested technological opportunity has waned, possibly as a result of declining funding for public sector agricultural R&D. A related controversy is the extent to which *future* productivity growth will slow in response to the decline in public sector funding for agricultural R&D since 2004.

Figures 3-5 can all be read as evidence supporting the idea that modern research has been less useful for new R&D than in the past. We express this in two different ways in this paper. First, we note that patents are increasingly turning to older ideas. Second (and largely equivalently), we document that ideas published in the 1970s and 1980s (and in some sectors, the 1990s) were cited more heavily, relative to an expected value, than ideas published since that time.

7. Conclusion

This paper has developed a new dataset to study how previously discovered knowledge is used in agricultural R&D. Using patents as instantiations of new commercially valuable agricultural inputs, we interpret the citations patents make as signals of knowledge flows in six different agricultural subsectors over 1976-2016.

We document a number of stylized facts. Agricultural patents tend to cite older ideas than non-agricultural patents, primarily due to the persistence of older ideas. This trend has become more pronounced over the last few decades. Put another way, the ideas generated in the 1970s and

1980s are more heavily cited, after controlling for various trends in patenting, than ideas generated in the 2000s and 2010s. This suggests old ideas are particularly valuable in agriculture. We obtain similar findings, in miniature for individual patents. Agricultural patents are more valuable and more likely to be cited if the age distribution of their ideas has a low mean but a high variance (meaning a long tail of older ideas). For cultivar patents, it may not even be true that a low mean outperforms a high mean. In general, however, we find it is hardware-based agricultural R&D, rather than biochemical agricultural R&D, that most heavily relies on older ideas.

That said, several aspects of these results would benefit from further study. Changes in patent law in the 1980s have likely skewed the age distribution of patents that rely on biological knowledge, as biological patents were only legalized in the 1980s. Understanding the extent to which this is a bias in certain biochemical subsectors would be useful. A second issue is the heterogeneity of non-patent references. It would be desirable to see if there are systematic differences in the age distribution of different kinds of non-patent references, such as citations to academic journal articles and citations to patent-related documents. Finally, to better inform the large literature linking agricultural R&D expenditures to agricultural productivity, further work is needed. Future work could develop lags between R&D *spending* and the publishing of cited documents, and also between the time a patent is granted and when innovations are adopted in the field. In this way, a more precise estimate of R&D lags that varies across sectors and time could be achieved.

This paper provides a new strand of evidence on an old claim: agriculture is peculiar in the extreme length required to fully realize the returns to R&D. A not insubstantial share of agricultural R&D is still cited decades after it is completed. Despite the acceleration of many things in recent decades, this story has not become less relevant. Indeed; the ideas used in agriculture are getting older.

References

Albers, Stevan C., Annabelle M. Berklund, and Gregory D. Graff. 2016. The Rise and Fall of Innovation in Biofuels. *Nature Biotechnology* 34(8): 814-821.

Alston, J.M., Norton, G.W., and Pardey, P.G. 1995. *Science under scarcity: principles and practice for agricultural research evaluation and priority setting*. Cornell University Press.

Andersen, M.A., and Song, W.. “The Economic Impact of Public Agricultural Research and Development in the United States.” Agricultural Economics 44,3(2013):287–95.

Arora, Ashis, Marco Ceccagnoli, and Wesley M. Cohen. 2008. R&D and the Patent Premium. *International Journal of Industrial Organization* 26: 1153-1179.

Baudry, Marc, and Béatrice Dumont. 2006. Patent Renewals as Options: Improving the Mechanism for Weeding Out Lousy Patents. Review of Industrial Organization 28(1): 41-62.

Bessen, James. 2008. The value of U.S. patents by owner and patent characteristics. Research Policy 37(5): 932-945.

Chavas, J., and Cox, T.. “A Nonparametric Analysis of the Influence of Research on Agricultural Productivity.” American Journal of Agricultural Economics 74,3(1992):583–91.

Clancy, M.S., and G. Moschini. 2017. Intellectual Property Rights and the Ascent of Proprietary Innovation in Agriculture. *Annual Review of Resource Economics* 9:53-74.

Clancy, Matthew. 2018. Inventing by combining pre-existing technologies: patent evidence on learning and fishing out. *Research Policy* 47(1): 252-265.

Fuglie, Keith, Matthew Clancy, Paul Heisey, and James MacDonald. 2017. Research, Productivity, and Output Growth in U.S. Agriculture. *Journal of Agricultural and Applied Economics* 49(4): 514-554.

Huffman, W., and Evenson, R.. Science for Agriculture: A Long-Term Perspective. 1st ed. Ames: Iowa State University Press, 1993.

Huffman, W.E., and R.E. Evenson. 2006. *Science for Agriculture: A Long-Term Perspective*. Hoboken, NJ: Blackwell Publishing.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty. 2000. The Meaning of Patent Citations: Report on the NBER-Case Western Reserve Survey of Patentees. *NBER Working Paper* 7631.

Jin, Y., and Huffman, W.E.. “Measuring Public Agricultural Research and Extension and Estimating Their Impacts on Agricultural Productivity: New Insights from U.S. Evidence.” Agricultural Economics 47,1(2016):15–31.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. Technological Innovation, Resource Allocation, and Growth. *The Quarterly Journal of Economics* 132(2): 665-712.

Lanjouw, Jean O., Ariel Pakes, and Jonathan Putnam. 1998. How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data. The Journal of Industrial Economics 46(4): 405-432.

Mukherjee, Satyan, Daniel M. Romero, Ben Jones, and Brian Uzzi. 2017. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances*, April 19.

Nagaoka, S., K. Motohashi, and A. Goto. 2010. Patent Statistics as an Innovation Indicator. In the *Handbook of the Economics of Innovation*, eds., B.H. Hall and N. Rosenberg. North Holland Publishing.

Pakes, Ariel, and Mark Schankerman. 1984. The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In R&D, Patents and Productivity, ed. Zvi Griliches. Chicago: University of Chicago Press for the NBER.

Pardey, P., B. Koo, J. Drew, J. Horwich, and C. Nottenburg. 2013. The evolving landscape of plant varietal rights in the United States, 1930-2008. *Nature Biotechnology* 31: 25-29.

Pray, C.E., and K.O. Fuglie. 2015. Agricultural Research by the Private Sector. *Annual Review of Resource Economics* 7:399-424.

Serrano, Carlos J. 2010. The dynamics of the transfer and renewal of patents. The RAND Journal of Economics 41(4): 686-708.

Wang, S.L., Ball, V.E., Fulginiti, L.E., and Plastina, A.. “Accounting for the Impact of Local and Spill-in Public Research, Extension and Roads on US Regional Agricultural Productivity, 1980–2004.” Productivity Growth in Agriculture: An International Perspective. Fuglie, K.O., Wang, S.L., and Ball, V.E., eds. Wallingford, UK: CAB International, 2012, pp. 13–32.

Wang, S.L., Heisey, P.W., Huffman, W.E., and Fuglie, K.O.. “Public R&D, Private R&D, and U.S. Agricultural Productivity Growth: Dynamic and Long-Run Relationships.” American Journal of Agricultural Economics 95,5(2013):1287–93.

Appendix

Figure A1. Distribution of patent citation ages by year (left) and cumulative (right)

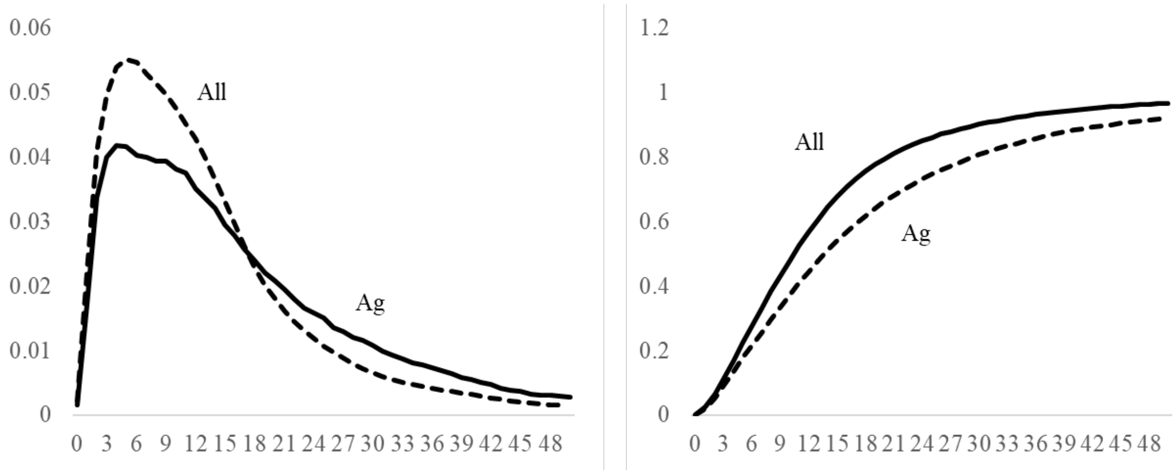

Figure A2. Distribution of non-patent reference ages by year (left) and cumulative (right)

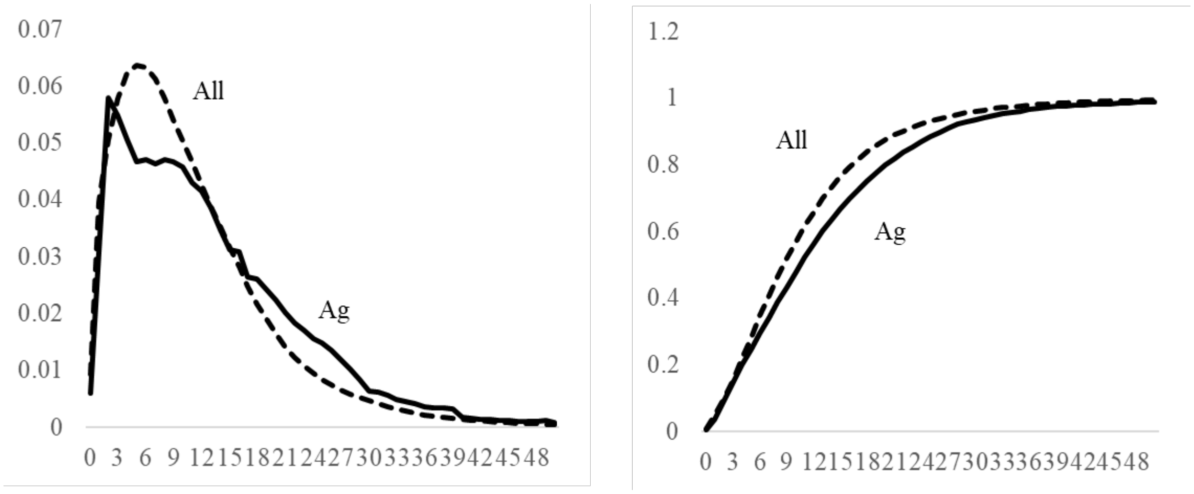

Figure A3. Actual relative to expected citations received by year research is published (patent citations)

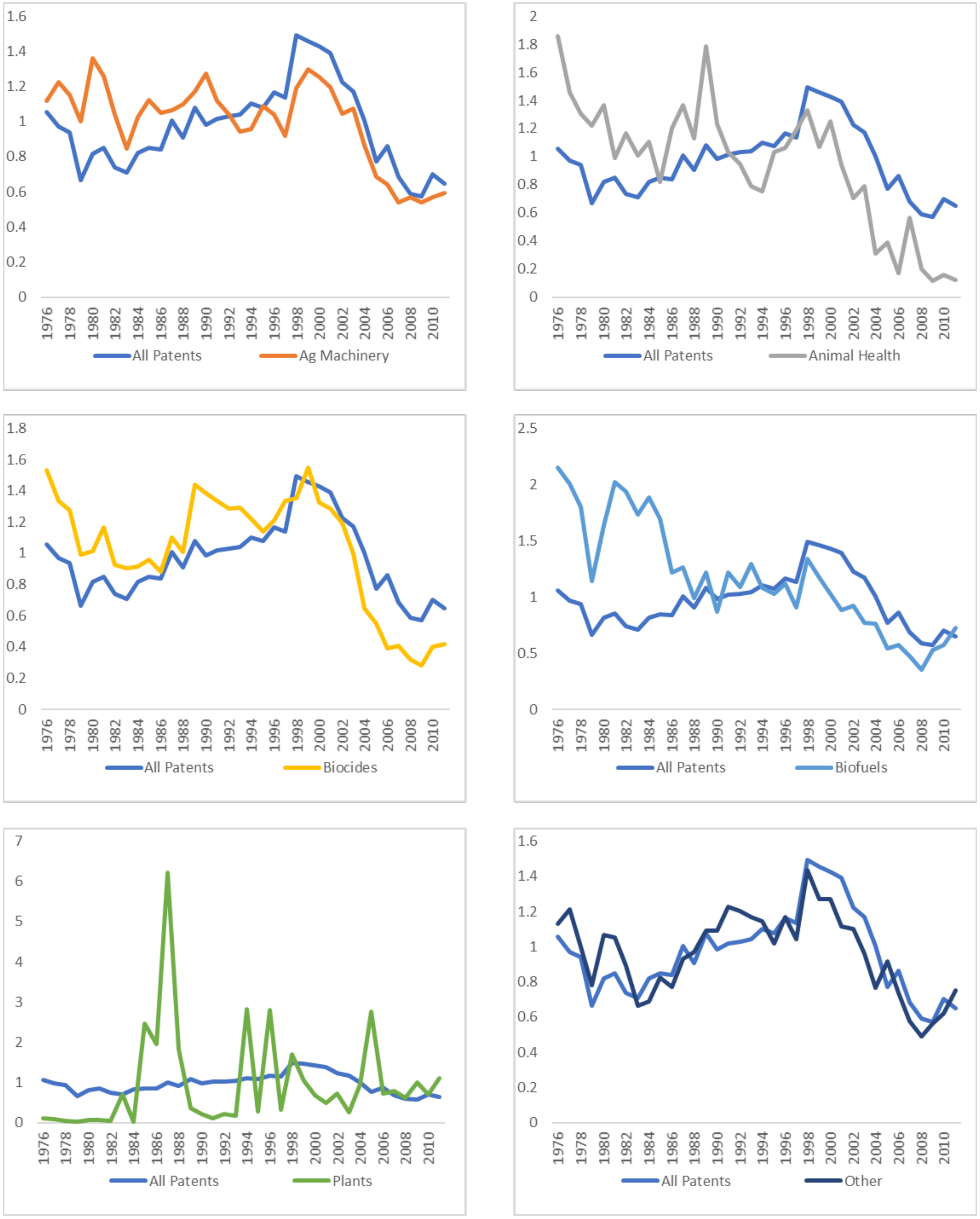

Figure A4. Actual relative to expected citations received by year research is published (non-patent references)

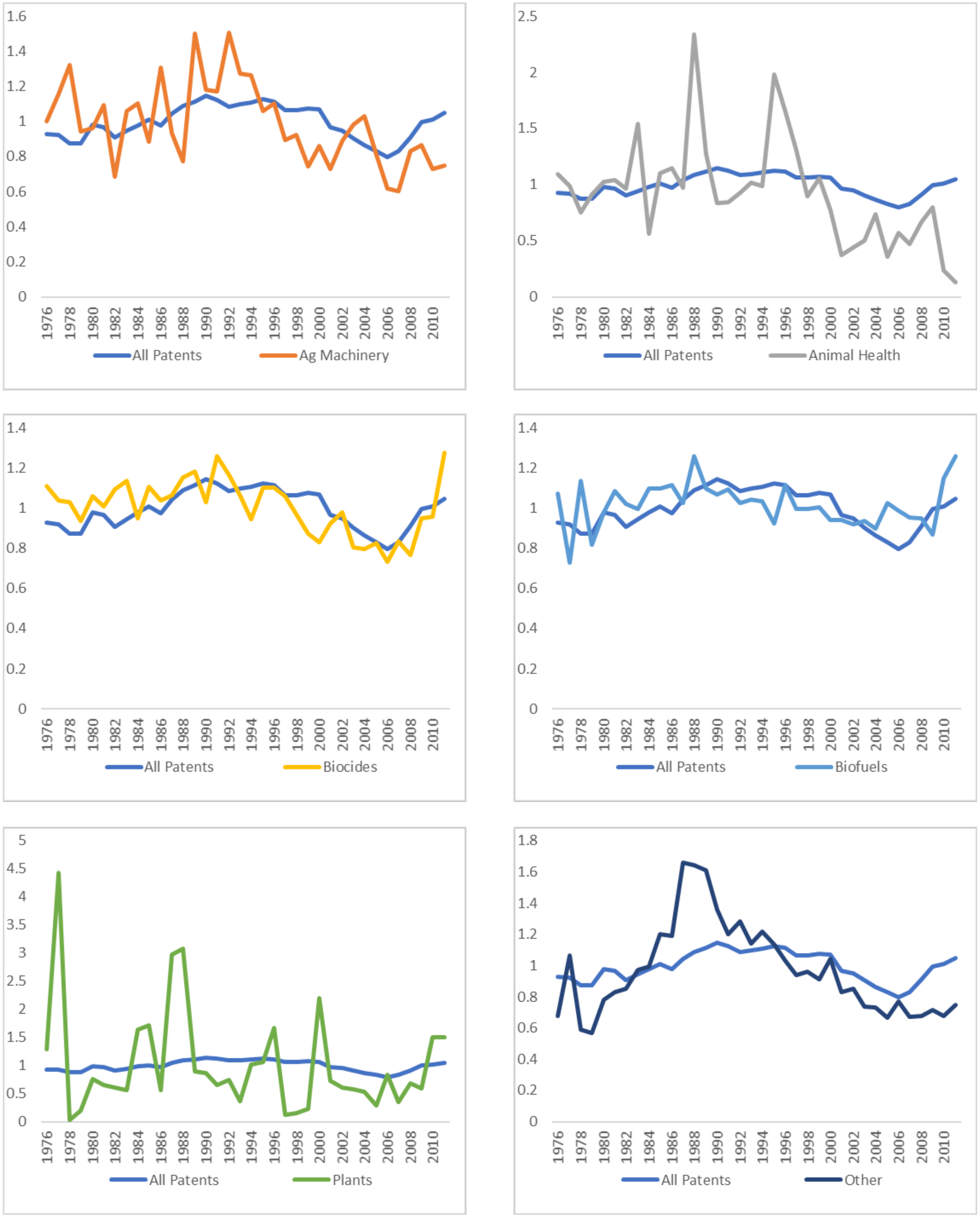

Table A1. Pooled patents robustness checks

Dependent Variable	**Renewal**	**KPSS**	**Citations**
Intercept	0.58***	-17.54	0.04
	(0.04)	(590.43)	(0.08)
Biocide	-0.41***	-2.26***	-0.79***
	(0.03)	(0.24)	(0.07)
Biofuel	-0.44***	-1.55***	0.15*
	(0.03)	(0.31)	(0.07)
Ag Machines	-0.39***	-2.94***	-0.36***
	(0.03)	(0.25)	(0.07)
Other Ag	-0.50***	-1.98***	-0.45***
	(0.03)	(0.26)	(0.07)
Plants	0.01	-1.80***	0.05
	(0.03)	(0.25)	(0.07)
IHS(cites made)	0.02***	-0.01	0.26***
	(0.00)	(0.06)	(0.01)
Publicly Traded	0.14***		0.21***
	(0.01)		(0.02)
Mean – patent citation	-0.00***	0.01*	-0.02***
	(0.00)	(0.01)	(0.00)
Mean – non-patent ref	0.00	-0.01	-0.00***
	(0.00)	(0.01)	(0.00)
C.V. – patent citation	-0.01	0.04	0.10***
	(0.01)	(0.10)	(0.02)
C.V. – non-patent ref	0.08***	0.13	0.27***
	(0.01)	(0.11)	(0.03)
Num. obs.	34,393	10,013	54,533

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$. Standard errors in parentheses. All regressions include year fixed effects

Renewal: linear probability model over probability patent is renewed to term (1982-2005)

KPSS: Logit model for probability Kogan et al. (2017) estimates among top 5% in sector (1976-2016)

Citations: Negative binomial regression over number of citations received in 5 years after grant (1976-2012)

Table A2. Ag subsector robustness checks

Dependent Variable	Animal Health	Biocides	Biofuels	Ag. Machinery	Plants	Other Ag
			Mean – Patent Citation Age			
Renewal	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	**-0.00** (0.00)	**0.01** (0.00)	**-0.00** (0.00)
KPSS	0.00 (0.09)	0.01 (0.01)	0.04 (0.07)	0.01 (0.01)	0.02 (0.04)	0.01 (0.02)
Citations	**-0.05** (0.02)	**-0.01** (0.00)	**-0.01** (0.00)	**-0.02** (0.00)	**0.03** (0.01)	**-0.02** (0.00)
			Mean – Non-Patent Reference Age			
Renewal	-0.00 (0.00)	**-0.00** (0.00)	-0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	**0.00** (0.00)
KPSS	-0.09 (0.10)	**-0.03** (0.01)	-0.05 (0.07)	0.00 (0.02)	0.01 (0.03)	0.00 (0.03)
Citations	0.01 (0.01)	**-0.01** (0.00)	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	**-0.01** (0.00)
			Coefficient of Variation – Patent Citation Age			
Renewal	-0.11 (0.10)	0.01 (0.01)	-0.09 (0.06)	0.03 (0.02)	0.08 (0.05)	**-0.04** (0.01)
KPSS	-1.15 (2.08)	0.00 (0.11)	0.32 (1.50)	0.17 (0.30)	-0.40 (1.52)	-0.12 (0.58)
Citations	0.25 (0.34)	0.03 (0.04)	0.22 (0.13)	0.07 (0.04)	**0.85** (0.17)	**0.12** (0.03)
			Coefficient of Variation – Non-Patent Reference Age			
Renewal	0.03 (0.10)	**0.03** (0.02)	**0.20** (0.07)	**0.19** (0.04)	-0.04 (0.04)	**0.14** (0.02)
KPSS	2.14 (2.78)	-0.29 (0.42)	-1.10 (1.66)	-0.19 (0.63)	**2.27** (0.76)	**1.59** (0.59)
Citations	0.37 (0.33)	**0.29** (0.04)	**0.64** (0.11)	**0.44** (0.08)	**-0.26** (0.13)	**0.12** (0.05)
			# of Observations			
Renewal	285	6768	909	10235	1659	14537
KPSS	201	3033	308	4051	1313	1107
Citations	392	9691	2187	16508	4220	21535

Notes: Bold coefficients indicates p<0.05. Standard errors in parentheses
All regressions include year fixed effects, log-hyberbolic sine of number of citations made, and a dummy if Kogan values are available.
Renewal: linear probability model over probability patent is renewed to term (1982-2005)
KPSS: Logit model for probability Kogan et al. (2017) estimates among top 5% in sector (1976-2016)
Citations: Negative binomial regression over number of citations received in 5 years after grant (1976-2012)