RAND Journal of Economics
Vol. 50, No. 3, Fall 2019
pp. 591–614

Endogenous productivity of demand-induced R&D: evidence from pharmaceuticals

Kyle Myers*

and

Mark Pauly**

We examine trends in the productivity of the pharmaceutical sector over the past three decades. Motivated by Ricardo's insight that productivity and rents are endogenous to demand when inputs are scarce, we examine the industry's aggregate Research and Development (R&D) production function. Using exogenous demand shocks to instrument investments, we find that demand growth can explain a large portion of R&D growth. Returns to scale have been stable, whereas total factor productivity has declined significantly. Predicted rents based on our estimates and Ricardo's theory closely match the trends we observe.

1. Introduction

■ Considerable debate surrounds the growing research and development costs for new drugs. These discussions often revolve around estimates of the average costs to bring a new drug to market, which are estimated to have increased nearly six-fold in real terms from the 1980s to the 2000s (DiMasi et al., 1991; DiMasi, Grabowski, and Hansen, 2016).[1] Most discussions take this trend as *prima facie* evidence of a growing friction in the R&D process, then conjecture as to why it has occurred and how it may be corrected (e.g., Ruffolo, 2006; Pammolli, Magazzini, and Riccaboni, 2011; Scannell et al., 2012).

Some blame regulatory burdens from purportedly stricter Food and Drug Administration (FDA) requirements. Some blame an overreliance on large-scale compound screening methods as opposed to "rational" investigations based on theory and hypothesis testing. Others blame health insurers for somehow pressuring drug firms to pursue products with lower success probabilities

* Harvard University; kmyers@hbs.edu.
** University of Pennsylvania; pauly@wharton.upenn.edu.

The Pharmaceutical Research and Manufacturers of America provided data that enabled our analyses, but no other support, financial or otherwise. We are also thankful for the thoughtful comments and suggestions from Joe DiMasi, Pierre Dubois, Tom Philipson, Neeraj Sood, and other attendees at the Bates White Life Science Symposium (2016) and the Becker Friedman Health Sector and the Economy conference (2017). All errors are our own.

[1] See, for example, Hewitt, Campbell, and Cacciotti (2011). These R&D figures are frequently cited during debates on pharmaceutical prices, although rarely following standard economic logic—*ex post* optimal prices are independent of *ex ante* sunk costs.

but higher health improvements. Last, some argue that the stock of easy to develop ideas—the "low-hanging fruit"—has shrunk and not been replaced. In response to these discussions, industry leaders have called for the development of new R&D models, and for the government to bridge the so-called "valley of death" in the drug discovery process (Butler, 2008; Collins, 2011).

More broadly, the declining pace of innovation and growing costs of R&D per output have been documented across the US economy (Jones, 1995; Gordon, 2012; Bloom et al., 2018). Still, it remains unclear to what extent these trends are driven by real economic frictions that might necessitate policy intervention, or are instead the expected outcomes of rational firms making investments in more costly, but increasingly demanded ideas.[2]

In this article, we clarify the nature of the productivity decline in pharmaceutical R&D, and offer evidence that this decline is consistent with theoretical predictions about the way productivity evolves when demand grows faster than the supply of ideas for new products. In short, our findings are most in line with the "low-hanging fruit" hypothesis of scarcity.

We connect the classic idea of Schmookler (1966), expounded by Acemoglu and Linn (2004)—the rate of innovation is directly related to demand growth—with Ricardo's (1817) point—demand and productivity will be inversely related when inputs (here, profitable new drug ideas) are rare. This connection guides our investigation of a simple aggregate R&D production function,

$$N = \alpha R^{\beta}, \tag{1}$$

where the number of new products is a function of R&D investments per productivity parameters α (TFP) and β (output elasticity). In our main analyses, we utilize the fact that firms' optimal investment level depends on the future size of the market, and identify the productivity parameters using Acemoglu and Linn's (2004, henceforth, AL04) exogenous demographics-driven measure of demand to instrument.

The pharmaceutical sector provides both an inherently important and empirically ideal setting to study R&D dynamics because we can (i) identify exogenous demand shocks, (ii) connect these shocks to R&D investments and new products, and (iii) separately identify changes to TFP and output elasticities. To do this, we utilize therapeutic-class-specific data on US consumer drug expenditures, private US-based R&D investments, and approvals of New Molecular Entities (NMEs, our proxy for new, highly valuable products).

In the sense that the exogenous demand measure is an instrumental variable, the reduced-form evidence is clear; the elasticity of new drugs with respect to market size, the focal parameter of AL04, is very stable over time. This is good evidence that any productivity decline was likely not driven by allocative inefficiencies at firms. If managers had somehow gotten worse at directing investments, we would have expected the relative rate of NMEs approved in larger versus smaller markets to decline.

However, due to changes in industry reporting, we lack disaggregated private investments post-2000, and cannot instrument investments throughout our sample (1985–2013). To overcome this limitation, we predict private R&D when unobserved under the assumption that firms are equally responsive to demand shocks over time.[3]

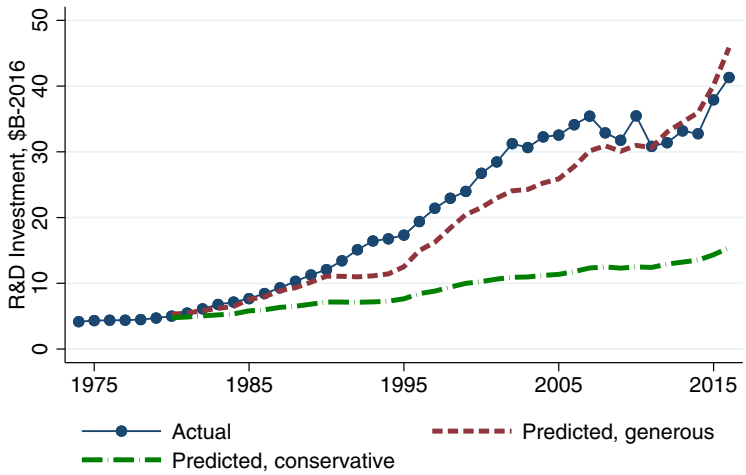
Figure 1 shows the nature of the predictive R&D model, described in detail below; it plots the sum of the predicted investments alongside actual investments to illustrate the role of expected market size in stimulating private R&D. Notably, demand alone can explain anywhere from about a third to virtually the entirety of the growth in total R&D since 1980.

[2] For instance, when discussing semiconductors and growing costs of sustaining Moore's Law, Bloom et al. (2018) note that it may be the case that "Demand for better computer chips is growing so fast that it is worth suffering the declines in idea Total Factor Productivity (TFP) there in order to achieve the gains associated with Moore's Law."

[3] This has some reasonable implications, for example, that managers didn't get better or worse at forecasting demand, but also implies a more debatable feature, that managers did not forecast any productivity shocks. We discuss the implication of this assumption in Section 3.

FIGURE 1

ACTUAL AND PREDICTED R&D TRENDS [Color figure can be viewed at wileyonlinelibrary.com]



Note: PhRMA-reported total domestic R&D spending on NMEs. Predicted values are based on pre-2001 disaggregated data and the demand-R&D investment equation (equation (3)). The generous model does not include time-period fixed effects when estimating equation (3), whereas the conservative model does, thus identifying the demand-investment relationship only based on within-class, within-time-period variation in demand. In both cases, correlation of actual and predicted investments is approximately 0.97.

Using these predicted values, we estimate a version of equation (1) over the full sample. In summary, our estimates fail to reject a null hypothesis that output elasticity (β) is stable—the industry's returns to scale haven't changed. However, we do identify significant declines in TFP (α). Together, the magnitudes imply that a growing market (per drug class) in 2016 will still induce the same, proportional increase in new drugs as in 1980. However, compared to 1980's averages, all 2016 markets received only about two thirds as many new drugs.

When combined with our robustness tests and reduced-form results, which show no significant changes to the new drug-market size elasticity, we are confident that these results are driven by the data and not any specific assumptions of our approach. We also explore the potential role of research investments by the National Institutes of Health (NIH). NIH investments do not appear to precede our demand shocks, supporting our exclusion restriction. However, the lack of correlation here is a part of our story—because we don't find evidence that the NIH's supply of research (which likely plays major role in the industry's TFP) is endogenous to demand, whereas private investments are, we should expect TFP declines following demand growth.

Finally, we discuss our results more generally in the context of Ricardo's theory of demand-driven productivity and rents. Suggestive evidence is very much in line with his two major predictions: during periods of demand growth (i) firms, facing scarce inputs, invest in ideas that are less "fertile"; and (ii) earn larger rents as a function of these declines in fertility. This raises doubts that there is any efficiency or management problem in the industry that needs to be corrected by public intervention or private exhortation. Of course, subsidies would still lower private sector costs of R&D and potentially improve TFP, but those subsidies may not be addressing a true market failure. Our explanation does not perfectly refute any of the alternatives.[4] It is a market-based explanation of why firms increasing chose to pick the "high-hanging fruit" even if the flow of new ideas has not diminished. In other words, it appears the long-run R&D supply curve in

[4] However, we comment in conclusion on some results which cast doubt on some of the alternative stories, and consider whether remedies based on them might do more harm than good.

this industry slopes upward. Our most normative result is to question the conventional wisdom that a decline in R&D productivity is a clear sign of misallocations by firms.

Since at least Schmookler (1966), economists have investigated the "pull" force of demand on innovation. The aforementioned AL04 provides a useful theoretical model of this notion supported with empirics that we expand on here. A number of other studies have investigated the elasticity of pharmaceutical innovations with respect to changes in demand, consistently finding significant positive effects.[5]

Our analysis also complements recent examinations of the productivity slowdown. In closely related work, Bloom et al. (2018) ask if "ideas are getting harder to find" and provide a clear discussion of R&D productivity in the context of macroeconomic endogenous growth models. Examining their "idea production function," which is loosely analogous to equation (1), they find declining R&D productivity within many sectors, including semiconductors, agriculture, and (using similar data as us) pharmaceuticals. This article is a useful complement to theirs in that: (i) Bloom et al. (2018) cannot make the policy-relevant distinction between α or β, whereas we can, and (ii) our consideration of the impact of demand growth as causing higher expected cost ideas to become profitable provides an important insight into one of the less discussed determinants of productivity decline, an explanation they can only hint at.

The remainder of the article is as follows: Section 2 outlines the theory of demand-driven productivity; Sections 3 and 4 describe the empirical approach and data, respectively; Section 5 presents the main results and robustness checks; Section 6 discusses the findings as they pertain to Ricardo's predictions and concludes.

2. Theoretical motivation

■ One of the major contributions of the economist David Ricardo was the distinction between price mechanisms for different kinds of production inputs. Ricardo argued the productivity of agriculture depends on the type of input. Most inputs (e.g., seed, animals) have their prices determined by the marginal cost of producing them. However, the price of land (rent) is a different story; in this case, rather than price being "cost determined," it was more correct to say that the price per input is determined by the price of the final product, which itself is a function of demand. The simple notion is that the stock of land with a given fertility is exogenous (or at least fixed in the short run), but its price is determined by demand. His key insight for our purposes is that, as increases in demand bring more, but less fertile, land into production, productivity on a per unit basis will necessarily decline. That decline is an inevitable result of moving on to the set of next most profitable opportunities in a market where the supply curve slopes up. Indeed, in this model with supply of opportunities fixed and growing demand, increases in demand must result in declining measured productivity. Furthermore, Ricardo notes that, given the scarce resources, these increases in demand will raise rents to land with high fertility; we consider the implications of this proposition later.[6]

We propose a similar process determining the R&D costs per new drug. Our implied theoretical model is as follows: each period, exogenous inputs, namely, demographic-driven demand growth, determine the distribution of new drug ideas per their expected profitability. An increase in demand will increase the return on investment across all potential drug opportunities, causing ideas with costs too high to justify investment under prior (lower) demand to become

[5] Finkelstein (2004) explores a policy change related to the value of vaccines and finds a 1% increase in expected market size stimulates a 2.5% increase in the number of clinical trials for affected diseases. Duggan and Morton (2010) and Blume-Kohout and Sood (2013) utilize Medicare Part D as a plausibly exogenous shock to consumers' willingness to pay for and also find corresponding increases in clinical trials for drug categories expected to grow the largest. Dubois et al. (2015) utilize detailed global revenue data and instrumental variables approach to estimate the NME-demand elasticity directly.

[6] The lack of any disaggregate revenues prevents an in-depth analysis, but in Section 6, we discuss a proxy for these rents—revenues per previously discovered and on-patent drug per market size—which displays a trend in line with Ricardo's prediction.

potentially profitable. However, importantly, unless these demand shocks are accompanied by equal growth in the supply of ideas, these newly profitable ideas still have the same lower expected productivity (higher costs) as when demand was lower.

That is, to the extent that production ideas are scarce, increasing the output of new drugs in response to demand has to lower the productivity of R&D. Unless there is an equivalent offset in the number or quality of ideas—perhaps driven by a surge in basic-science investments—increasing R&D costs per new product is expected.

In the context of the production function described by equation (1), what are "ideas" and why might they be scarce? Following Jones (2005), we conceptualize ideas as instructions for converting inputs (R) into outputs (new drugs); changes in the distribution of ideas will manifest as changes in productivity parameters. Traditionally, ideas or knowledge are referenced as a chief component of what we identify as TFP (i.e., a Hicks-neutral production shift), with output elasticities held fixed. However, in the context of R&D, such an assumption warrants testing because the totality of ideas is (obviously) unknown. Thus, we remain agnostic as to how a scarcity of ideas would influence either of the productivity parameters.

Practically speaking, what drives differences in the costs associated with these ideas and why might this change over time? Looking across ideas, Budish, Roin, and Williams (2015) highlights the importance of variation in the costs across projects for different types of cancers and find evidence of distortions away from research on early-stage cancers, which have a larger costs due to the length of clinical trial necessary.[7] As the worldwide burden of diseases continues to shift from one driven by acute to chronic diseases (Vos et al., 2015), the relative demand for treatments for these complicated, long-term conditions has certainly grown. In turn, this means that R&D projects must be designed to handle more complex outcomes over longer periods of time. In line with this argument, Ward et al. (2015) report a near doubling in clinical development time for new antiviral drugs,[8] and Getz and Campo (2017) document anywhere from a 20%–70% increase in the complexity of clinical trials based on the number of procedures and visits in trial protocols.

More generally, Jones (2009) provides theoretical and empirical evidence from patents that the cumulative nature of knowledge requires increasingly specialized (and implicitly costly) efforts over time to generate the same level of advances. Simply put, because all new knowledge must somehow build on prior knowledge, and that prior knowledge must first be learned by the inventor, it will become increasingly costly to push the frontier forward.[9] To summarize, we do not argue that ideas of *any* kind become increasingly scarce over time, but rather, in agreement with Jones (2009) and Bloom et al. (2018), that ideas of equal expected benefit (in terms of potential revenues) are becoming more costly to pursue. In other words, the supply curve becomes increasingly steep (upward sloping) as the hunt for new ideas continues.

3. Empirical approach

■ For the production function, we model the number of NMEs approved by the FDA in therapeutic class j in year $y = (1, 2, \ldots, Y)$ as a standard conditional Poisson model,

$$\mathrm{NME}_{jy} = \frac{\exp\left(\alpha_{t(y)} + \beta_{t(y)} \log(R_{jy})\right)}{\sum_{v=1}^{Y} \exp\left(\alpha_{t(v)} + \beta_{t(v)} \log(R_{jv})\right)} + \epsilon_{jy}, \tag{2}$$

[7] Because the clinical trials must follow patients for much longer to reach clinical endpoints. The authors also discuss the role of "short termism" whereby agency problems in the management process induces inefficient discounting that incentivizes the pursuit of shorter clinical trials (Budish, Roin, and Williams, 2015).

[8] Although they can only investigate new antivirals approved in the United Kingdom, they also report no significant changes in the time spent during regulatory approval.

[9] Assuming that the marginal costs of education remain fairly constant.

where years are grouped into four-year time periods t and R_{jt} is prior private R&D investments, as described below.[10] This grouping of years into time periods facilitates our estimation of the parameters, given the relatively small number of observations within each year, and mimics AL04's original analyses. The Poisson specification accommodates the count nature of our dependent variable and focuses on the rate of innovation.[11] We condition out time-invariant differences across therapeutic classes because these differences appear to be substantial; they account for roughly 25% of the observed variation in NME output and nearly 70% of the observed variation in private R&D investments across classes.[12] Note also that equation (2) does not directly identify TFP levels as written in equation (1), but instead, the α_t parameters are estimates of relative (percentage) changes in TFP over time.

Because private firms do not fund R&D projects at random and we cannot identify exogenous firm characteristics, estimating equation (2) without considering the demand for certain types of drugs would confound firm-level and endogenous market-level characteristics, biasing our estimates. Therefore, we consider a two-stage model where the level of R&D investments in class j at time $t - l$ is determined by expected market size in (then future) year y per the investment function

$$\log(R_{jy}) = \gamma_0 + \gamma_D \mathbb{E}[\text{Market}]_{jy} + \delta_j + \tau_{t(y)} + \varepsilon_{jy}, \tag{3}$$

which also conditions on fixed cross-sectional differences (δ_j) and time trends ($\tau_{t(y)}$). Following changes in demand, firms adjust R&D investments per equation (3) which in turn determines NME output per equation (2). Our motivation and process for lag (l) selection is described in Section 4 below.

When disaggregated data on private spending is available, we can estimate a joint variant of the investment and production equations using instrumental variables via Generalized Method of Moments(GMM), as outlined by Blundell, Griffith, and Windmeijer (2002).[13] Given our data limitations of lacking post-2000 category-specific investments (described below), we will focus largely on a two-step approach to estimating these equations under the assumption that the causal effect of market conditions on investment decisions estimated in the first stage (γ_D) is persistent. Then, substituting predicted private R&D into the second equation for the full sample, we can test our main hypotheses about changes in α and β over time.

Notably, the method of two-stage-predictor-substitution (2SPS) in nonlinear models has been shown to potentially suffer from bias (Terza, Basu, and Rathouz, 2008). Ideally, we would jointly estimate the equations as a single nonlinear instrumental variables equation or using two-stage-residual-inclusion. However, given the data constraints, neither are possible over the full sample. We address this concern in two ways. First, using years when we can estimate equations (2) and (3) jointly, we compare the coefficients generated by joint estimation and 2SPS to approximate the magnitude of bias. Second, we also estimate linear versions of our equations. In both cases, the absolute magnitudes and, more importantly, relative trends in parameters across specifications is consistent. In terms of our theory, fixing γ_D implies that managers cannot predict

[10] A limitation of this model is that it prevents any spillovers across therapeutic classes. To consider how this could influence our analysis, consider the case where positive spillovers did exist and the entirety of the declines in productivity were due to declines therein. In our specification, these changes would be captured by a declining and conflate the traditional notion of TFP with this particular source of marginal returns. However, we note that many of the traditional policy levers discussed to spur TFP are very similar to those that would influence the degree of spillovers across research lines—develop infrastructure, coordinate overlapping investments, and promote information sharing.

[11] We explore other estimation models also designed to handle count data: negative binomial and zero-inflated Poisson regressions.

[12] These estimate are from log-transformed OLS models with therapeutic-category fixed effects, because these fixed effects are not estimated in the conditional Poisson specification. Although these linear specifications are obviously biased, they provide a sense of magnitudes.

[13] We kindly thank Timothy Simcoe for making a Stata implementation of this procedure available on his website at: people.bu.edu/tsimcoe/data.html.

TABLE 1 Summary Statistics

	All Avail. (1)	1980s (2)	1990s (3)	2000s (4)	2010s (5)
Panel A: Drug-Class Level, 65 groups; 1980–2016					
NME Approvals	0.371	0.258	0.446	0.354	0.448
	(0.847)	(0.572)	(0.826)	(0.737)	(1.244)
Expected Market Size ($B)	2.490	1.656	2.200	2.959	3.426
	(3.765)	(2.365)	(3.152)	(4.247)	(4.977)
Expected Market: Patients (M)	32.14	25.95	29.71	34.97	40.41
	(45.19)	(35.71)	(40.80)	(48.27)	(56.01)
Expected Market: Per-person $	138.5	117.5	134.7	154.3	151.6
	(218.5)	(184.8)	(210.8)	(242.6)	(234.6)
Obs.	2405	650	650	650	455
Panel B: PhRMA R&D Level, eight groups; 1980–2000					
Private R&D ($B)	1.506	0.845	2.048		
	(1.384)	(0.655)	(1.579)		
Obs.	168	80	88[A]		
Panel C: NIH Level, 16 groups; 1980–2016					
Total Grants ($B)	0.813	0.469	0.658	1.130	1.072
	(0.791)	(0.298)	(0.526)	(1.003)	(0.975)
Obs.	592	160	160	160	112

Note: Mean values within groups (at the Anatomical Therapeutic Chemical (ATC-3), Pharmaceutical Research and Manufacturers of America (PhRMA)-reported, or National Institutes of Health (NIH)-based level) and year, s.d. in parentheses. Research and Development (R&D) and NIH grants deflated to 2016-$ per the Biomedical Research and Development Price Index (BRDPI); market-size measures deflated to 2016-$ per the Consumer price index (CPI). [A]Includes 2000.

any productivity declines or improvements. Given the multiyear lags between project initiation and eventual success or failure, this implication is reasonable.

Because we are estimating a production function and will be arguing that our demand shocks are orthogonal to the productivity parameters, our estimates are relevant for investments spurred by any market force or regulation that is also orthogonal to productivity. Many policies that influence this sector are likely to meet this criteria: for example, health insurance coverage, patent duration or stringency, or R&D tax subsidies.[14]

4. Data and variable construction

■ This section describes our main data sources and how they were utilized to construct our three key variables: New Molecular Entities (NMEs) approved by the FDA, private pharmaceutical R&D, and exogenous expected market size. Table 1 summarizes these and other variables used in the analyses.

□ **Drug classes and FDA approvals.** Determining the specificity of drug classes for our empirical analysis involves a trade-off between the ability to control for important fixed differences across classes (e.g., stable underlying scientific potential or components of demand) versus allowing for spillovers, given the idiosyncratic nature of R&D. Although private R&D investments are only available for eight drug categories as detailed below, the demand measure can be decomposed into more specific categories, because drug-level data is available. As a compromise, we use the Anatomical Therapeutic Chemical (ATC) Classification System as a guide to match the eight industry R&D categories to 65 corresponding subgroups of the ATC hierarchy—the three-digit "ATC-3" codes—as shown in Appendix A.

[14] So long as these policies are not themselves endogenously motivated by productivity shocks.

The real outcome we are concerned with is welfare changes as a result of new products released by the pharmaceutical industry. Because data necessary to calculate welfare is rarely available (i.e., drug-specific revenues and quality-adjusted life years generated), it has become common practice for studies similar to this to evaluate the count of new drugs approved. Obviously, a raw count of all drug approvals will place the same weight on all drugs, whether they are revolutionary therapies such as statins or reformulations of age-old drugs, such as aspirin. To alleviate some of this discrepancy, it has also become common practice to restrict attention to approvals at the FDA that receive NME status. This status indicates that the active moiety has not yet been approved by the FDA, thus providing a strong indication that the therapy has potential to provide significant welfare improvements. Thus, throughout the article, we are forced to assume that NMEs are homogeneous.[15]

The count of NME approvals from 1987 to 2016 was constructed from the Drugs@FDA database. The database does provide information as to the sponsoring firm of each drug approval; however, it is very difficult to identify the "ownership" of any given drug over time, given the prevalence of firm- and drug-level acquisitions, as well as licensing and manufacturing agreements. Thus, we do not restrict our sample to any set of firm sponsors. Drugs were matched to each therapeutic category per their assigned ATC codes.

□ **Private R&D expenditures.** Estimates of industry-wide investments in pharmaceutical R&D from 1970 to 2016 were constructed based on annual reports from the Pharmaceutical Research and Manufacturers of American (PhRMA), the industry's lead trade group. PhRMA conducts annual surveys of its member companies and reports summarized results for a number of relevant statistics. Using historical reports,[16] we obtained PhRMA-wide estimates of: total industry R&D, the share of R&D allocated to new/innovative product lines (e.g., NMEs), and the share of R&D allocated across eight major therapeutic categories, as outlined in the Appendix. The share of R&D allocated to NME-type research is not decomposed by therapeutic categories and is reported in only a select number of annual reports, 1998–2000. The average reported allocation is roughly 80%, and so we scale total investments by this amount.

The main caveat to this data is that, although we observe total R&D for all years, therapeutic category-specific investments are not available post-2000. Hence, the projection methodology described in the proceeding section. In the Appendix, we compare the industry-wide totals from these reports to those from the National Science Foundation (NSF's) survey of R&D investments, and find the PhRMA data to very closely match the NSF data, with no significant discrepancies.[17]

Our final estimate of industry investment must account for the notoriously long development times in this industry—investments in any given year may be related to drugs anywhere from 1 to sometimes 20 years away from final approval. Thus, relating changes in R&D investments to same-year NME approvals may not reflect the true connection between the two. Instead of making *ad hoc* assumptions about lag periods, we take a data-driven approach and conduct multiple regressions of NME approvals on prior R&D investments for lag times between 5 and 10 years (given average reported development times of about 10 years in the PhRMA reports), and choose the lag from the model with the largest R^2. The Appendix reports the results of these regressions, from which we choose a 6-year lag, and test the robustness of our final results also to 9-and 12-year lags, given the average total development time of about 10 years, reported in the PhRMA data.

All private investments are deflated using the Biomedical Research and Development Price Index (BRDPI), because we are interested in the effective productivity of the industry's inputs

[15] A Deloitte LLP report found that average forecast peak one-year sales for NMEs is roughly $400M–900M, about 2.25 times greater than the same metric for non-NMEs (Deloitte, 2015).

[16] Reports prior to 2002 are not publicly available, but PhRMA representatives kindly shared copies of annual reports from 1990 to 2001, containing data from as early as 1972.

[17] Our results will not be biased by any overreporting or underreporting, so long as firms are not differentially doing so in both surveys over time. We have no reason to believe otherwise.

and not nominal investments. The BRDPI is developed by the NIH to track price de-/inflation specific to goods and services that are employed in biomedical research.[18]

□ **Exogenous potential market size.** In their initial analyses of market size and innovation in the pharmaceutical sector, AL04 develop a plausibly exogenous measure of market size. This measure utilizes demographic trends to remove the influence of innovation on demand (consumers buying more new products because they are valued) in order to only identify the influence of demand on innovation (firms developing new products because they expect them to be valued by consumers).

The exogenous expected market size for drug category j in year y is given by

$$\mathbb{E}[\text{Market}]_{jy} = \sum_{a} I_{ay} \times S_{aj}, \tag{4}$$

where a is a set of five-year age bins, I_{ay} is the aggregate national income of individuals in age bin a year y, and S_{aj} is the average share of group a's income spent on drug category j over the course of the data. Because of the sparse Medical Expenditure Panel Survey (MEPS) data for many drug classes, we also follow AL04 in weighting all regressions by the standardized number of observations used to generate the expenditure shares.[19]

AL04 provide evidence that within-group drug expenditure shares are relatively constant over time, although the country's demographics are not. Intuitively, illnesses and the medications used to treat them often affect humans at certain ages (e.g., very few people under 45 take statins, but this share increases dramatically with age), so as the income of individuals of certain ages grows, so does the demand for drugs that are differentially utilized by their age group. Following AL04, the income component is constructed from the Current Population Survey (CPS) March supplement, and the expenditure component is constructed from the Prescribed Medicine Files of the Medical Expenditure Panel Survey (MEPS).

Drug categories are determined based on matching drug names reported in MEPS to their ATC. The MEPS data is only available from 1996 to 2016, therefore, we extend the average calculated during this time frame to our full time frame of 1985 to 2016.[20] CPS data is available for the full length of our study, with all data deflated using the Consumer Price Index.

5. Main productivity results

□ **Reduced form: new drugs and market size.** We begin by estimating a reduced-form model, where the rate of new drug approvals is a function of demand as implied by the two-stage NME-R&D-Demand relationship discussed earlier. These regressions mimic the main specifications of AL04 and amount to estimating equation (2) but using the exogenous demand measure as the dependent variable. Table 2 presents the results, phasing in drug-class fixed effects (Col. 2) and time-period fixed effects (Col. 3), and then allowing the coefficient on market size to vary pre-/post-2000 (Cols. 4–6).[21] Columns 4–6 suggests that the NME-demand elasticity did not change after 2000, where a 1% increase in expected market size induced a 1.7% increase in NME approvals. Figure 2 Panel (a) recreates an even more flexible version of Col. 4, revealing

[18] See: officeofbudget.od.nih.gov/gbipriceindexes.html for more. We prefer this index over the CPI for deflating industry investments, as it is designed specifically for biomedical research. Using broader measures of inflation, namely the CPI, return very similar results, which are available upon request.

[19] As in AL04, the unweighted estimates (not shown) are very similar to the weighted, albeit with larger standard errors. See AL04 for a further discussion of this choice, noting their reliance on weights for all main specifications.

[20] We use all of the available MEPS data to most accurately identify stable averages of S_{aj} that are most likely to be driven by exogenous, latent expenditure shares due simply to the aging process.

[21] When including time controls, our point estimates are roughly half of AL04's. See AL04 Table 2, Panel C, Column 1, where they report a NME-demand elasticity of 3.54 (S.E. = 1.19). The difference is likely due to the differences in the years of data coverage and drug-class aggregation: their data examines NMEs from 1970–2000 using only MEPS from 1996–1998, and use only 16 major drug-class categories.

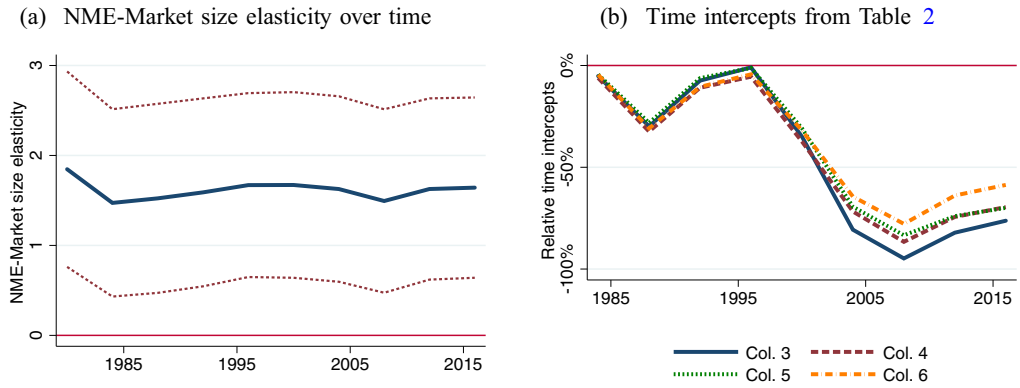
TABLE 2 Reduced-Form NME Production Function, 1980–2016

	NME_{jy}					
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{E}[Market]_{jy}$	0.342***	0.525**	1.628***			
	(0.0228)	(0.217)	(0.506)			
$\mathbb{E}[Market]_{jy}$				1.737***	1.611**	1.628**
$\times$ Pre-2000_y				(0.528)	(0.707)	(0.721)
$\mathbb{E}[Market]_{jy}$				1.729***	1.605**	1.619**
$\times$ Post-2000_y				(0.526)	(0.705)	(0.719)
Obs.	2405	2405	2405	2405	2405	2405
Drug-class FE		Y	Y	Y	Y	Y
Time FE			Y	Y	Y	Y
Spec.	Poisson	Poisson	Poisson	Poisson	Neg-Bin	ZIP

Note: Standard errors in parentheses, clustered at R&D-class level. * p<0.10, ** p<0.05, *** p<0.01. Drug-class Fixed Effects (FE) reflect the conditional Poisson model. "Neg-Bin" and "ZIP" are Negative Binomial and Zero-Inflated Poisson models, respectively.

FIGURE 2

REDUCED-FORM NME PRODUCTION OVER TIME [Color figure can be viewed at wileyonlinelibrary.com]



Note: Based on reduced-form regressions of New Molecular Entity (NME) approval rates on expected market size. Panel (a) recreates Col. (4) from Table 2, allowing the elasticity to vary more flexibly over time with dashed lines indicating 95% Confidence Interval (C.I.), whereas Panel (b) plots the time-period intercepts from the regressions in Cols. (3–6) from Table 2.

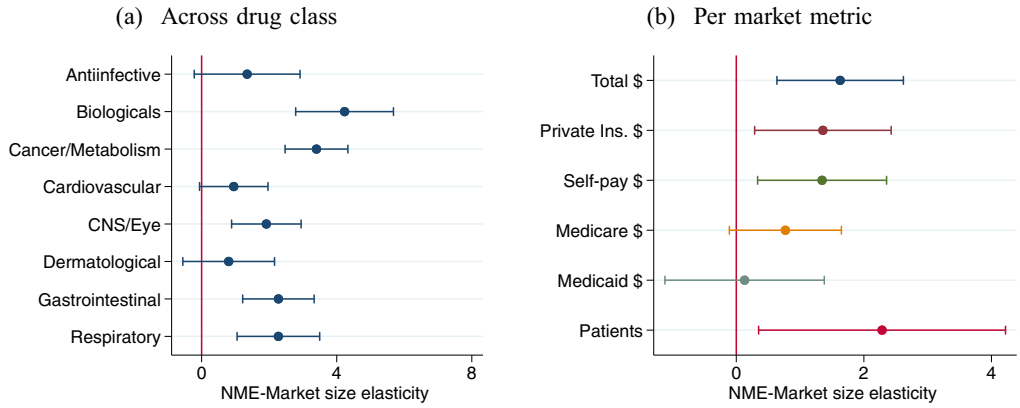
no notable trend in the NME-demand elasticity over this period, and suggesting that declines in the industry's output elasticity (β) are unlikely.

However, the average rate of NMEs across all classes declined substantially, as evidenced by the significant drops in the time-specific intercepts in these reduced-form regressions (see Figure 2 Panel b). This gives initial evidence that, in terms of equation (1), any productivity changes were likely concentrated within the TFP term (α).

In order to explore heterogeneity in the industry's responsiveness to different types of demand shocks, Figure 3 plots estimates of the NME-demand elasticity when allowed to vary across the eight major PhRMA R&D classes (Panel a) and using a variety of alternative expenditure categories when calculating expected market size (Panel b). Panel (b) indicates that there is some heterogeneity in the NME-demand elasticity across therapeutic classes ranging from about 1 to 4, with Cancer/Metabolism and Biologicals exhibiting the largest elasticities. Panel (b) indicates that if we define market sizes based only on expenditures from private insurers or self-payments, the elasticity is still estimated at roughly 1.5. However, we find smaller elasticities when using

FIGURE 3

HETEROGENEITY IN THE ELASTICITY OF NEW DRUGS WITH RESPECT TO MARKET SIZE [Color figure can be viewed at wileyonlinelibrary.com]



Note: Based on reduced-form regressions of New Molecular Entity (NME) approval rates on expected market size, with brackets indicating 95% Confidence Interval (C.I.). Panel (a) is based on a single regression, where the demand-NME relationship is allowed to vary over time or major drug class. Panel (b) presents results from six separate regressions, where the expected market size metric is based on either total pharmaceutical spending (Total $, the preferred specification), spending only from private insurance, self-paid, Medicare, or Medicaid dollars, or the number of population-based measure.

only Medicare- or Medicaid-specific expenditures to build the expected market size metric. This indicates that firms are less responsive to shifts in demand driven by these public insurance programs relative to private insurance coverage or the ability of patients to pay out of pocket. Panel (b) also includes results from when we only include population counts (and not their income) when constructing the expected market size measure (a robustness also explored in AL04).[22] The identified elasticity has larger standard errors than the other specifications, which is to be expected, as not including income data forces us to treat equal-sized populations equivalent, even if one may be wealthier and thus more willing to consume a particular medication class.

□ **First stage: R&D investments and market size.** Table 3, Columns (1–2) presents the first-stage results, where we regress prior private R&D on current values of the expected market size measure per equation ((4)) for the restricted portion of our sample, where we observe disaggregated private investments (1980–2000). As expected, we identify a significant positive relationship where an 12% increase in R&D spending is made in six years in advance of a 10% exogenous growth in market size. As expected, given the results in Figure 2, Column (2) shows no significant panel variation in this elasticity. The stability of this relationship over time lends support to our 2SPS approach in the following sections.

To get a sense of how much of the total growth in R&D spending this relationship might explain, we project class-specific R&D investments post-2000 and collapse these to an industry predicted total to compare with actual spending totals, which are observable over the full sample. For the purposes of this exercise, we assume no lag between the demand measure and private investments, so that we may project values for the full sample. Figure 1 plots the predicted and actual investments over time, where our "conservative" specification includes time-period fixed effects in the first-stage equation, whereas the "generous" specification does not. The figure indicates that somewhere between a third and virtually all of the growth in annual R&D spending since 1980 could plausible be linked to demand growth. We also note that one feature of demand in

[22] In this metric, aggregate income (I_{ay}) is replaced simply with the number of individuals in each age bin.

TABLE 3 Investment and Productivity Results, 1980–2000

	Private $R\&D_{j(y-6)}$		NME_{jy}		
	(1)	(2)	(3)	(4)	(5)
$\mathbb{E}[\text{Market}]_{jy}$	1.283***				
	(0.186)				
$\mathbb{E}[\text{Market}]_{jy}$		1.227***			
$\times$ Pre-1990$_y$		(0.186)			
$\mathbb{E}[\text{Market}]_{jy}$		1.264***			
$\times$ Post-1990$_y$		(0.189)			
Private $R\&D_{j(y-6)}$			0.736***	1.075**	1.398***
			(0.186)	(0.498)	(0.521)
Obs.	1365	1365	1365	1365	1365
Drug-class and Time FE	Y	Y	Y	Y	Y
First-stage	Y	Y			
IV				Y	Y
Spec.	OLS (log-log)	OLS (log-log)	Poisson	GMM Poisson	2SPS
Implied Investment Response: $\frac{\$R\&D}{\mathbb{E}[\$\text{Demand}]}$	0.670				
Implied Marginal Cost: $\frac{\$R\&D}{NME}$ (M)			432.0	295.5	227.3

Note: Standard errors in parentheses, clustered at R&D-class level. Columns (1–2) provide first-stage estimates of equation (3), Columns (3–5) provide estimates of equation (2) using either endogenous (Col. 3), instrumented (Col. 4), or predicted investments (Col. 5).* p<0.10, ** p<0.05, *** p<0.01. F-statistics for the demand proxy in the first-stage investment regressions are as follows: 47.7 (Col. 1), 43.4 (Col. 2, pre-1990), 44.7 (Col 2., post-1990). Investment Response reports the amount of R&D funds spurred by one additional dollar of expected market size. Marginal costs are per NME in millions. "Time FE" in Cols. (3–5) are the time-varying relative TFP estimates, because these include the production function.

this sector that the exogenous market size measure does not accommodate is an income elasticity of health different than 1. In order to maintain exogeneity, the measure implies that a constant share of income is spent on health. Although empirical evidence is mixed (Acemoglu, Finkelstein, and Notowidigdo, 2013), there is good reason to suspect this elasticity is above 1, given the dual investment and consumption nature of health goods (Hall and Jones, 2007). Thus, our (untestable) hypothesis is that this measure is underpredicting real demand growth.

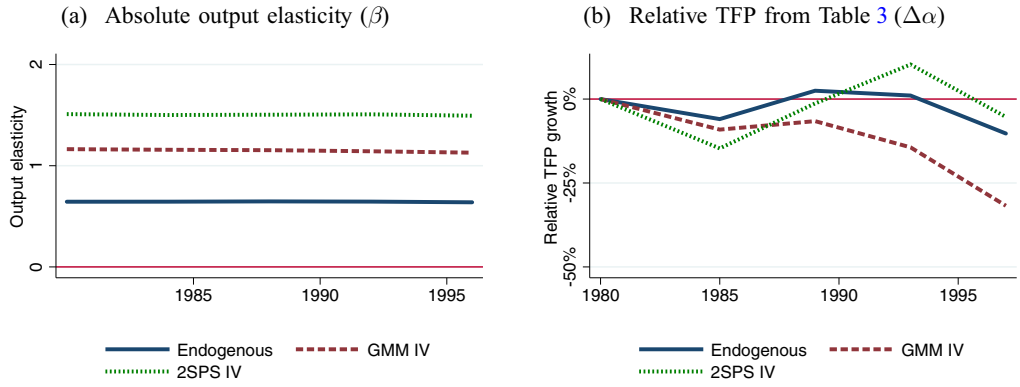
□ **Two stage: R&D productivity.** Before investigating productivity over the full sample, we first must validate our 2SPS approach within the pre-2000 period. To do so, Table 3 present our estimates of for the period when we observe disaggregate private R&D investments, 1980–2000. Estimates are based on raw Poisson, regressions (Col. 3), GMM Poisson, where we estimate equations (3) and (4) together (Col. 4), and the 2SPS Poisson, where we use predicted R&D investments from the endogenous investment equation (4) to estimate the production function equation (3) (Col. 5). When using the exogenous demand measure to instrument R&D spending, we identify an output elasticity of about 1 (i.e., constant returns to scale), which implies an average cost per marginal NME of roughly \$295M.[23]

However, to reiterate, our focus is the evolution of these costs over time. Also, because joint estimation of the investment and production functions (as in Cols. 5–6) is not feasible over the full sample, we are forced to utilize the 2SPS procedure to estimate the equations post-2000. As noted earlier, this methodology has the potential to produce biased estimates. To first get a sense of the magnitude of this bias, the 2SPS estimate is about 25% larger than the GMM estimate. However, we are only concerned with whether or not this bias changes over time, and so Figure 4 plots time-varying estimates of β and α in Panels (a) and (b), respectively. First off, we see virtually no

[23] For comparison, the accounting-based estimates of average costs for this time-period range from \$200M to \$500M (DiMasi et al., 1991; DiMasi, Hansen, and Grabowski, 2003). Our estimates are on the same order of magnitude, though, because the other estimates are based on average costs and proprietary project-level data from manufacturers, it is unclear to what extent any fixed costs or selective reporting may influence their estimates.

FIGURE 4

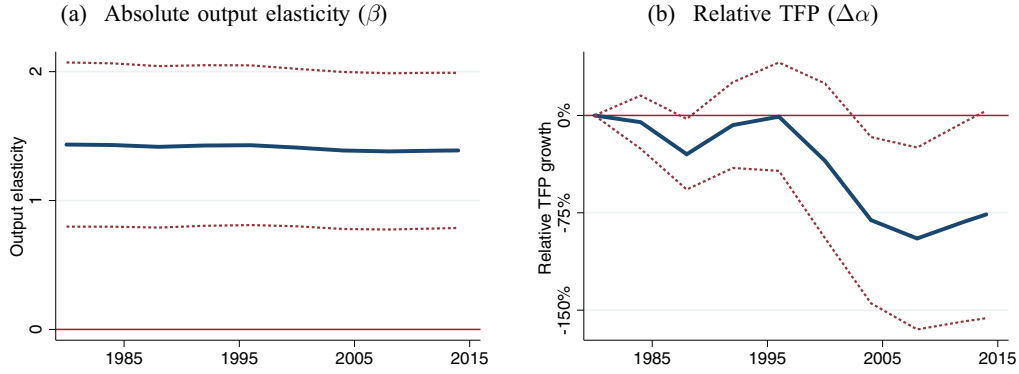
TIME-VARYING ESTIMATES OF PRODUCTIVITY PARAMETERS, 1980–2000 [Color figure can be viewed at wileyonlinelibrary.com]



Note: Panel (a) recreates Columns (3–5) from Table 3, allowing either output elasticity to vary over time. Panel (b) plots the TFP-estimates from directly from the regressions in Columns (3–5), connected at years starting each time period (i.e., value at 1997 corresponds to estimate for 1997–2000 period).

FIGURE 5

TIME-VARYING ESTIMATES OF PRODUCTIVITY PARAMETERS, 1980–2016 [Color figure can be viewed at wileyonlinelibrary.com]



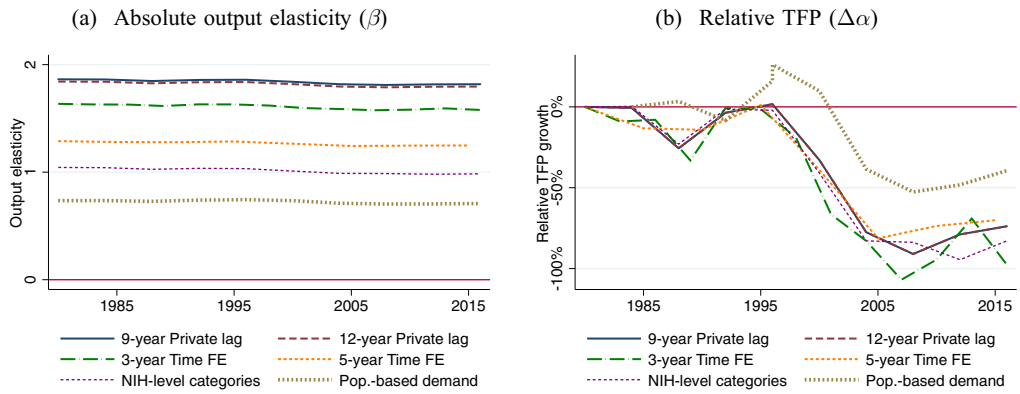
Note: Estimates are from two separate regressions based on the production function equation (2) with dashed lines indicating 95% C.I., where in Panel A, the output elasticity parameter (β) is allowed to vary over time periods, and in Panel B, the TFP parameter (α) is allowed to vary over time periods.

change in the output elasticity over time, regardless of estimation approach. In the case of TFP, we do find that the 2SPS approach may underestimate the apparent decline that is occurring in this period. Still, given the correlation in trends, especially in Panel (a), we are confident that any trends in coefficients that arise using the 2SPS approach over the full sample are data-driven and not artifacts of the specification.

Our main productivity results are presented in Figure 5. Panels (a) and (b) present the results of estimating the 2SPS version of the investment and production functions and allowing the coefficients, respectively, to vary over time periods. Virtually all of the changes in productivity arise within the α (TFP) coefficient, whereas the returns to scale β (output elasticity) of this industry are stable. This pattern is in line with the results presented in Figure 2 Panel (a), where the NME-demand elasticity (which depends only on β) is also very stable. However, we identify

FIGURE 6

ROBUSTNESS TESTS—TIME-VARYING PRODUCTIVITY PARAMETERS [Color figure can be viewed at wileyonlinelibrary.com]



Note: Recreates Figure 5 using six alternative specifications, where private R&D investment is lagged 9 or 12 years (instead of 6), the time-period fixed effects are based on 3- or 5- year increments (instead of 4), the data is collapsed to the 16-level "NIH"-based groups (instead of the 65 ATC-3-based groups), or the demand measure is based only on patient population sizes (instead of total spending).

declines in TFP on the order of 75%, which corresponds to all drug classes receiving roughly two thirds as many NME approvals in the later time periods.

Figure 6 recreates Figure 5 using six different alternative specifications to explore the sensitivity of the focal trends. First, we vary the lag between prior private R&D investments and the year of NME approval, testing 9- and 12-year alternatives. Next, we set the time-period fixed effects to span three or five years instead of four. Then, we collapse the data to drug-class groupings based on the NIH data (described below), which provides a slightly more aggregated data with 16 classes. Next, we estimate the model using the population version of the expected market size metric. In all of these cases but the population-based demand metric, we find patterns that are very similar to the preferred specification—output elasticities are very stable, and TFP declines by about 75%–100%. Given the heterogeneity in NME-demand elasticity across drug classes observed in Figure 3 Panel (a), Appendix B describes an additional version of the analyses, where we allow the investment response (γ) and output elasticity (β) to vary across these classes. The TFP changes over time identified allowing for this heterogeneity are approximately 100%, in line with the main results.

□ **Public supply-side: NIH investments in science.** Our key assumption underlying this analysis is that the demographics-driven demand shocks only affect firms' optimal R&D investments and not the productivity parameters, nor any other variable, that in turn influences the productivity of private R&D. The most obvious concern is that other funders of biomedical research may also expect these demand shocks, and endogenously respond in a manner that influences both firms' decisions and the ultimate rate of new products observed, which would bias our estimates. Following AL04, we focus on the US National Institutes of Health (NIH), and explore the possibility that the NIH and the scientists who seek funding there might also respond to the demand shocks.

There is some evidence that as a whole, the direction of basic biomedical research moves in advance of downstream disease burdens (Bhattacharya and Packalen, 2011). At the NIH in particular, there is evidence that private lobbying efforts—presumably motivated by demand—can influence the direction of public funds (Hegde and Sampat, 2015). Also notably, recent empirical evidence based on exogenous "windfalls" of funding at the NIH identifies significant positive

downstream effects of NIH-funded research on private patenting rates—even when focusing only on the patents of drug candidates (Azoulay et al., 2018). Although Azoulay et al.'s (2018) results provide strong support for the notion that firms respond positively to marginal investments by the NIH—the government can "push" supply—we are concerned here with whether or not the government (and the scientists seeking funding) are themselves "pulled" by demand alongside private firms. If this is the case, it presents an econometric concern for our analyses in that we could not separate to what extent firms are responding specifically to the market, or to the NIH, which itself is responding to the market. If this is not the case, we can be confident that the induced R&D we observe is driven solely by the market; however, this raises a potential policy concern because, given the findings of Azoulay et al. (2018), it may be more efficient if both private and public investments are induced by demand growth.

To examine these relationships, we utilize data from the NIH and estimate (i) correlations between NIH funding and the demand shocks that would suggest an endogenous public response, (ii) correlations between NIH funding and the demand-induced private R&D investments that would also suggest endogeneity, and (iii) an additive production function to recover and compare the marginal productivity of public and private R&D.[24]

Investments via NIH extramural grant awards from 1965 to 2016 was constructed based on data available from the NIH ExPORTER data files as well as a Freedom of Information Act request for grant dollar amount data prior to 2000 that is not readily available in the public files. In line with prior research on the allocation and impact of NIH funds (Toole, 2012), we use the funding institutes and keywords in the grants' titles to group NIH spending into 16 classes.[25] As with private R&D, all NIH investments are deflated using the BRDPI to ensure comparison of effective inputs.

Table 4 reports the regression estimates for the three relationships of interest: (Panel a) OLS estimates of NIH funding regressed on the exogenous market measure, (Panel b) OLS estimates of NIH funding regressed on predicted (per demand) private R&D, and (Panel c) conditional Poisson estimates of NME counts regressed on predicted private R&D and observed NIH funding. To allow for flexible timing sequences, we evaluate these relationships using NIH investments in the 6, 8, 10, 12, 14, or 16 years prior to the focal year of NME approvals. Across this range of lags, Panels (a) and (b) reveal no significant correlation between NIH investment levels and either demand shocks or the private investments they induce. This provides evidence that the induced private R&D investments we observe are not motivated in part by simultaneous increases in productivity driven by NIH investments.

Panel (c) of Table 4 estimates a version of the NME production function (equation 2) now including NIH funding. The relationship between NIH funding and NME output appears somewhat sensitive to the lag specification, with coefficients ranging from about 0.2 to 0.5 (i.e., a 10% increase in NIH funding in year $y - l$ is correlated with a 5% increase in NME output in year y). The coefficient on predicted private R&D is consistent with estimates not including NIH funding, again in line with the lack of correlation between NIH and private R&D observed in Panel (b).

The marginal product of these public investments can be estimated under the assumption that the year-to-year variation in NIH funding (conditional on the drug-class fixed effects) is driven by variation uncorrelated with other factors that may directly influence NME output (e.g., idiosyncratic political preferences that influence NIH budgets, or the "windfalls" examined by Azoulay et al., 2018). Based on the average coefficients for each source of investments, the estimates suggest that one "NME's worth" of additional private R&D investments ($270 million) is equivalent to about $1 billion in NIH funding. Although, we caution interpretation of these

[24] This final estimation of the additive production function is similar to the analyses in Toole (2012); however, instead of including the AL04 demand instrument and observed private R&D directly in the production function, we include the demand-instrumented private R&D variable.

[25] See the Appendix for further details on the NIH categorization approach.

TABLE 4 Supply-Side—Role of the NIH

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Role of Demand						
$NIH_{j(y-l)}$; OLS						
$\mathbb{E}[Market]_{jy}$	0.415	0.324	0.113	0.330	−0.141	0.250
	(0.273)	(0.237)	(0.236)	(0.256)	(0.318)	(0.323)
Panel B: Role of Demand-Induced Private R&D						
$NIH_{j(y-l)}$; OLS						
$\widehat{Private R\&D}_{j(y-6)}$	0.323	0.252	0.0880	0.257	−0.110	0.195
	(0.341)	(0.299)	(0.277)	(0.318)	(0.341)	(0.417)
Panel C: Private + Public NME Production Function						
NME_{jy}; Poisson						
$\widehat{Private R\&D}_{j(y-6)}$	1.016***	1.104***	1.175***	1.165***	1.335***	1.168***
	(0.371)	(0.374)	(0.362)	(0.373)	(0.379)	(0.376)
$NIH_{j(y-l)}$	0.498***	0.460***	0.524***	0.459***	0.207*	0.176**
	(0.123)	(0.116)	(0.125)	(0.118)	(0.110)	(0.0830)
Obs.	2405	2405	2405	2405	2405	2304
NIH lag (l)	6	8	10	12	14	16
Implied Private MC: $\frac{\$R\&D}{NME}$ (M)	298.6	274.6	258.3	260.7	227.8	259.9
Implied NIH MC: $\frac{\$Grant}{NME}$ (M)	977.0	978.3	798.9	835.7	1674	1808

Note: Standard errors in parentheses, clustered at R&D-class level. $\widehat{Private R\&D}_{j(y-6)}$ denotes the predictions of private R&D per the exogenous demand measure. All models include time fixed effects and either includes drug-class fixed effects directly (OLS) or conditions out these effects (Poisson). All variables except for NME counts are log-transformed. * p<0.10, ** p<0.05, *** p<0.01.

results within the context of this study, and note that our purpose here is only to address the potential concern that other major supply-side forces may be responding to the demand shocks we identify and confounding our private R&D productivity estimates. To that effect, we are confident our results are not being significantly influenced by such changes. From an econometric standpoint, this alleviates certain endogeneity concerns. However, as we discuss below, from a policy standpoint, this may be a source of friction.

6. Discussion

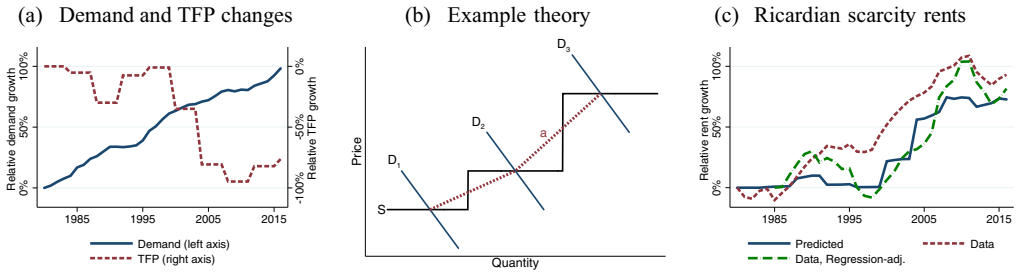
□ **Revisiting Ricardo's theory.** To summarize our findings at this point, we identify significant declines in TFP for the industry since 1980, whereas returns to scale (output elasticity) have remained stable. Returning to Ricardo's notion of scarce resources, we now investigate the extent to which his theory explains patterns of this industry. Ideally, for the following analysis, we would identify many market-time-specific estimates of demand, productivity, and rents. Then, we could systematically relate demand shocks to changes in productivity and examine the connection to rents. With our data (only nine estimates of TFP changes) this is not possible, and in general, would necessitate an unusually large and detailed data set. However, given the usefulness of understanding the extent to which Ricardo's theory underlies R&D dynamics, we believe it is worthwhile to investigate the predictions.

The first straightforward prediction of Ricardo's model is that, to the extent some inputs are scarce, demand growth will lead to productivity declines. In terms of our data and results, this implies a negative correlation between proportional changes in demand and TFP.[26] Figure 7 Panel (a) plots the relative aggregate demand growth across all drug classes against the estimated relative changes in TFP (recreated from Figure 5 Panel (b)). Aggregate demand increased steadily throughout the time frame, except for a slight pause in the early 1990s, essentially doubling from 1980 to 2016. The TFP estimates exhibit an equal but opposite-signed change over this period,

[26] The degree of the correlation will be driven by the scarcity of inputs.

FIGURE 7

RELATING DEMAND GROWTH TO PRODUCTIVITY AND RENTS [Color figure can be viewed at wileyonlinelibrary.com]



Note: Displays growth in measures relative to baseline estimates or data from the 1980–1983 period. In Panel (a), the demand shifts are based on data, whereas the TFP shifts are recreated from Figure 5. The text walks through the simple theory of scarcity rents using Panel (b). Panel (c) plots the predicted scarcity rents based on the results of Panel (a), and compares them to two proxies described in the text.

with the majority of the decline occurring post-1995. The correlation (though based only on nine observations) is approximately -0.85, in line with the argument that larger demand growth will lead to larger productivity declines.

The second prediction of Ricardo's theory is that these demand-driven productivity declines will lead to corresponding increases in the rents accrued to owners of the scarce inputs. Also specifically, the magnitude of these rents will be based on the product of the size of the demand shocks and the productivity changes. To clarify, consider the simple supply-demand graph in Figure 7 Panel (b). This is a static representation of our dynamic analysis. Mimicking our results, this supply curve has been drawn to have constant returns to scale locally around the three equilibria defined by demand curves $D_{\{1,2,3\}}$ to reflect our point estimates of β that are generally near 1.[27] However, drawing S as a step function globally allows us to reflect our result of declining TFP over time—the line a represents the inverse of the TFP trend.

Firms receive scarcity rents depending on the evolution of TFP (the shape of S) and the size of demand shocks (e.g., the shift to D_2 from D_1). Larger TFP declines amid larger demand shocks will generate more rents. However, both changes must occur; the TFP declines must be spurred by a demand shock. If demand increases and TFP is constant, firms accrue no scarcity rents (by definition). Also, if there is no demand shock, then there is no movement along the supply curve. To rephrase this, changes in rents will be proportional to the changes in productivity multiplied by changes in demand: $\Delta\text{Rent} \propto \Delta\text{Productivity} \times \Delta\text{Demand}$.

Our goal is to use this relationship to predict rents given observed demand growth and estimated TFP declines, and then compare this prediction to what we find in the data. Due to the manner in which we estimate TFP over time—in terms of growth relative to the first time period of the sample—we construct our measures of demand and rents to also describe relative growth since the earliest data points (with the prefix $\%\Delta$).[28] Our measure of relative demand growth is simply the average percent change in the exogenous potential market size measure across all drug classes. Again, variation in this measure is driven only by demographic trends of the Unites States over this time period. Combining this with our TFP estimates yields our rent predictions: $(-1 \times \%\Delta\text{TFP}_y) \times \%\Delta\mathbb{E}[\text{Market}]_y$.

Using the data to directly construct rents in the Ricardian sense is less straightforward. Ideally, a measure of these rents would relate time-varying drug-specific prices to the original

[27] The estimates in Figure 5 Panel (a), which are likely overestimates in absolute terms due to the 2SPS approach, all include an elasticity of 1 in their confidence intervals, and the robustness tests in Figure 6 Panel (a) are all centered around 1.

[28] We use the years 1980–1983 to define the baseline period against which all future periods are compared.

average R&D costs required to bring those same drugs to market. Because we can only observe total annual NME approvals and sales, we take two approaches to estimating realized rents. First, we construct a data-based rent proxy per

$$\%\Delta\left[\frac{\text{Revenue}_y}{\mathbb{E}[\text{Market}]_y} \Big/ \sum_{y'=y-10}^{y} \text{NME}_{y'}\right], \tag{5}$$

where for each year y, we divide the ratio of total revenues per potential market size by the number of "on-patent" NMEs. Intuitively, scaling revenues by our exogenous demand measure—which approximates consumers' total willingness to pay for new drugs—reflects the share of consumer surplus that firms are able to capture. Scaling the measure again by the number of drugs approved in the prior 10 years ensures two things: (i) that we are comparing inframarginal drugs developed during similar time frames and likely have similar development costs, and (ii) account for the number of new patent-protected products on the market at a given time.[29] Under the assumption that both the marginal quality of NMEs and the marginal costs of production (conditional on discovery) are stable over time, variation in this measure will only be driven by price effects, and therefore rent, as desired.

In our second approach, we estimate excess returns as a function of market size and NME count, using the following regression:

$$\text{Revenue}_y = \eta_1 \times \mathbb{E}[\text{Market}]_y + \eta_2 \times \sum_{y'=y-10}^{y} \text{NME}_{y'} + \xi_y. \tag{6}$$

We then define our regression-adjusted estimates of rent using $\%\Delta\xi_y$. These residuals describe excess returns beyond what we would have expected, given the average revenue associated with different levels of market size (per η_1) and the industry's NME portfolio (per η_2).[30]

Figure 7 Panel (c) plots relative changes in our predicted and estimated rents over time. In terms of relative changes, the two data-based estimates and the predicted growth in industry rents track very closely. Rents are relatively stable from 1980 to 1995, after which there is a significant increase until about 2010, after which rents appear to have stabilized. Compared to the early 1980s, the industry in the late 2000s was obtaining 75% larger revenues relative to the costs of developing those drugs.[31]

Overall, these trends are in line with what Ricardo would expect if "good" and "cheap" new drug ideas are scarce—the long-run supply curve slopes upward, or more colloquially, there is no more "low-hanging fruit." However, we note that these patterns are only preliminary evidence consistent with Ricardo's model.[32]

This suggestive result that the changes in rents accrued to this industry could be explained by a theory of scarce inputs has important implications for policies focused on drug prices. Traditionally, policy makers intent on lowering drug prices have discussed mechanisms such as implementing cost-effectiveness thresholds (at least, outside the United States) or promoting consumer bargaining power (e.g., by preventing consolidation or putting political pressure on firms). Our results imply another, indirect mechanism: subsidize investments in TFP. By lowering development costs, policies could enable some number of currently unprofitable ideas (e.g., due

[29] Based on a true patent length of 20 years and 10 years from discovery to final approval, the latter based on estimates reported in PhRMA reports.

[30] The difference between these residuals and the direct proxy first described is that they do not assume equally proportional effects of larger markets or a larger on-patent NME portfolio (i.e., the data proxy assumes that $\eta_1 = \eta_2$).

[31] In Appendix B, we present alternative lag formulations instead of the 10-year window used here, and still find the data-based rent proxies to align with the prediction.

[32] Recall that we may be underestimating the productivity declines due to the 2SPS IV approach. This would imply that our predicted rents would be larger, suggesting that, if anything, it is likely that industry rents did not grow as fast as one would have expected. This could be due to increasing competition within the industry itself, or a decline in these firms' bargaining power within the pharmaceutical supply chain.

to a low expected price) to become profitable investments in expectation. However, the addition of new products, even at lower prices, could still lead to increased total spending.

The TFP declines we estimate, and the rents they give rise to, behave just as Ricardo predicted they would in an efficient but input-constrained market. Still, what exactly is this "TFP" and how might policies influence it? As discussed earlier, we interpret this parameter as being driven by the amount of known but yet-to-be-exploited knowledge that the industry can access at a given time. Or more practically speaking, the number and quality of new drug targets generated by investments in basic science at institutions such as the NIH.

To this effect, the NIH recently established the National Center for Advancing Translational Sciences in order to facilitate the commercialization of ideas through efforts such as applied research projects and clinical trial support networks (Collins, 2011). Although, recall Section 5 and the lack of a relationship between NIH investments and future demand. This suggests that there may be gains to be had from directing the scientific workforce to pursue problems that are most likely to be demanded in the future, especially given evidence that exogenous funding shocks at the NIH do in fact lead to new drug developments (Azoulay et al., 2018).

□ **Conclusion.** Utilizing demographics-driven changes in demand to instrument R&D investments, this article estimates the productivity of the pharmaceutical sector over the past three and a half decades. Not only is this demand instrument useful for its econometric properties, but because the productivity estimates identified are based on investments made directly in response to changes in market size—a lever that policy makers often pull in hopes of stimulating R&D. Yet, since Ricardo, economists have noted that when such demand growth occurs in an industry where inputs are scarce, firms' productivity will necessarily decline in exchange for larger rewards in the market.

We argue that such a model provides an apt description of R&D, where the chief input of ideas is inherently in short supply. Based on our data from pharmaceuticals, we find that although output elasticity has remained very stable, TFP has declined significantly. In line with Ricardo's predictions, the periods of greatest TFP declines coincided with the greatest demand growth, which together led to increased rents.

Reconnecting with prior work examining this period of productivity in the pharmaceutical sector (e.g., Ruffolo, 2006; Pammolli, Magazzini, and Riccaboni, 2011; Scannell et al., 2012), most explanations to date have focused on some form of firm mismanagement, increasing regulatory burdens, or the lack of any more "low-hanging fruit." We cannot explicitly rule out the first two. As the large literature on across-country TFP differences has demonstrated, seemingly small misallocations of capital can have large effects on aggregate productivity measures (see Restuccia and Rogerson, 2013 for a review). The fact that the output elasticity of investments did not appear to change over this period casts some doubt on the role of misallocations at least *within* the pharmaceutical industry. Conditional on the inputs available, if managers became increasingly inefficient at allocating resources across drug discovery projects—due to new regulations or any other secular trend—we would have expected to see smaller returns to scale within each time period over time. Instead, larger pharmaceutical markets continue to receive the same larger proportion of new drugs that come to market; it is the *total* count of new drugs that has declined. If reallocation is relevant here, this would seem to suggest that is with respect to the allocation of resources across industries that share inputs with the pharmaceutical sector. Exploring this possibility seems to be an important avenue for future work.

However, given the important role demand plays with respect to driving investments and being correlated with productivity shocks, we believe our findings point mostly to the low-hanging fruit story, although not the one previously described. For instance, Pammolli, Magazzini, and Riccaboni (2011) use highly detailed project-level data on drug candidates to show that the classes of drugs that have experienced the largest growth in candidates pursued are also those with the lowest probabilities of success. They take these low success rates as given and argue that because those low-success markets have grown, productivity has declined. Our results indicate that lower

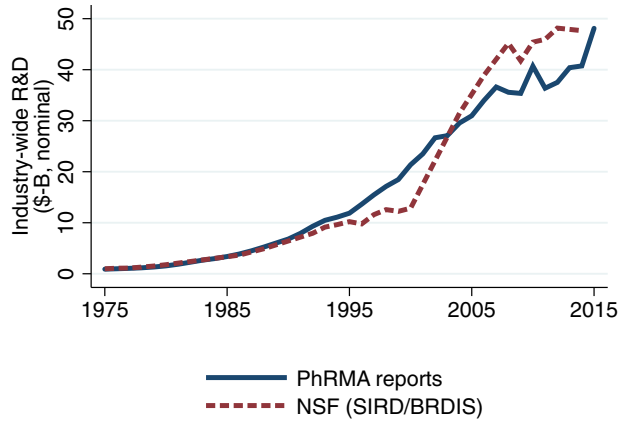
productivity can itself be a direct, endogenous result of market growth. We hope future research may shed more light on the nature of scarce ideas in the discovery process, such that R&D policies can more holistically address this interconnected nature of demand, productivity, and rents.

Appendix A

□ **R&D data comparison.**

FIGURE A1

PhRMA ANNUAL REPORTS VS. NSF SURVEYS [Color figure can be viewed at wileyonlinelibrary.com]



Note: Plots nominal industry-wide domestic R&D totals (in billions) using either the PhRMA Annual Reports, or the NSF's R&D survey data (Survey of Industrial Research and Development (SIRD) pre-2008, Business Research and Development and Innovation Survey (BRDIS) 2008-onward; per "Pharmaceuticals and Medicines" industry, North American Industry Classification System (NAICS) code: 3254).

□ **NIH categorization process.** Although the PhRMA and MEPS data is either already categorized or easily connected to existing data, the NIH data (especially pre-2000) lacks an immediately useful mutually exclusive categorization scheme. The following outlines our protocol for assigning NIH investments (per the total costs associated with grants) to the 16 categories outlined in Table A1, using a keyword-based approach common to studies of this industry (i.e., Toole, 2012). These 16 NIH-specific categories were developed through an iterative process where we tried to match the ATC-3 categories as close as possible, subject to constraints on how specific we could be, given the data. The data elements used included the title of the grant, the study section that reviewed the application, and the funding institute.

First, we retrieved only the "R," "P," "K," and "U" series of NIH grants for 1965–2016. These series typically account for about 85% of all grants. Data on grant total costs pre-2000 were obtained via Freedom of Information Act (FOIA). Restriction to these series was partly due to data availability and partly due to negotiations during the FOIA application to minimize the workload. This mostly excludes NIH investments in activities likely to have negligible impacts on drug development (i.e., community services program, loan repayments, and internal construction contracts), but also drops a few activities that may have nonnegligible effects (i.e., the "T" series training grants for pre- and post-doctoral students, or NIH intramural research projects). Still, given that the vast majority of funds flow through the four series we measure, we are confident that any variation in investments in these other minor activities is highly correlated with the R/P/K/U awards.

With grant costs in hand, we first rely on the study sections to classify grants. Using a keyword search (viewable in code provided), we classify each study section to one of the 16 major categories most likely to be relevant. However, due to very limited data in the older periods (i.e., many study section labels are missing or not linkable to any description due to churn), we are only able to reasonably classify about 35% of awards using this method. For this remainder, we search the grant titles using the same keywords as in the study section classification. We evenly divide each grant's costs among any category flagged using this approach. This leaves about 40% of awards still unclassified. For this final set, we construct Institute-Year share measures of how the total amount of classified investments (the other 60%) are split among the 16 categories. Then, using these shares, we divide the total amount spent on the unclassified grants accordingly (e.g., in a year, Institute i invested s% into category x, so we assign s% of Institute i's unclassified grant costs that year into category x).

□ **PhRMA—ATC—NIH crosswalk.**

TABLE A1 Category Crosswalk

PhRMA	ATC-3	NIH
Anti-Infectives	Anti-Fungals/Antibiotics (D01, D06)	Nonviral Anti-Infection
	Anti-Infectives (J01-02, J04)	
	Anti-Virals (J05)	Anti-Viral
	Anti-Parasitics (P01-03)	Parasites
Biologicals	Immuno-stimulants/suppressants (L03-04)	Immunomodulation
	Anti-Neoplastics (L01)	Cancer
	Anti-Diabetics (A10)	Diabetes/Obesity
Cancer/Endocrin./Metabolism	Corticosteroids (D07)	
	Hormones (H01-02, H05)	Endocrine—Hormones
	Endocrine Therapies (L02)	
	Sex Hormones	Endocrine—Reproduction
Cardiovascular	Anti-Thrombotics/-Hemorrhagics/-Anemics (B01-03, B05)	Blood
	Cardiovascular System (C01-C10)	Cardiovascular
Central Nervous Sys./Eye	Ophthalmologicals (S01)	Eye
	Nervous System (N01-07)	Neurological
Dermatological/Musculoskel.	Emollients (D02)	Musculoskel./Skin
	Anti-Psoriatics (D05)	
	Anti-Acne/Other (D10-11)	
	Musculoskel. (M01, M03-05)	
Gastrointestinal/Genitourinary	Alimentary Tract (A01-07, A11, A16)	Gastro-Intestinal
	Genito-Urinary (G01-02, G04)	Kidney, Gynecol., Urolog.
Respiratory	Respiratory System (R01, R03, R05-07)	Respiratory

Note: Private R&D categories are the most disaggregated levels of research and development investments reported in historical PhRMA reports. Anatomical Therapeutic Chemical Classification System (ATC) categories are used to classify drugs in the drug approval and utilization (demand) data. Each three-digit code is used as a separate category in the analyses. NIH classification schemes are described in further detail below.

□ **Determination of private investment lag.**

TABLE A2 Demand-Investment Relationship, Lag-Varying, Pre–2000

	PhRMA-Reported Investment$_{j(y-l)}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\mathbb{E}[\text{Market}]_{jy}$	1.557***	0.993***	1.158***	1.086***	0.942***	1.173***	1.539***
	(0.175)	(0.270)	(0.124)	(0.161)	(0.155)	(0.190)	(0.154)
Obs.	1300	1300	1300	1235	1170	1105	1040
Lag years (l)	4	5	6	7	8	9	10
R^2	0.836	0.844	0.864	0.857	0.836	0.831	0.814

Note: Standard errors in parentheses, clustered at R&D-class level. * p<0.10, ** p<0.05, *** p<0.01.

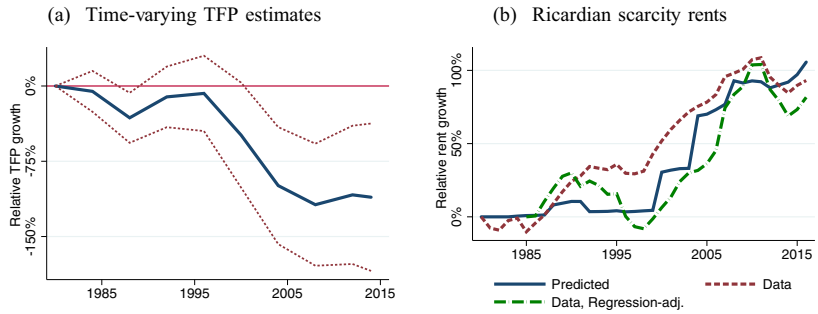
Appendix B

□ **Allowing drug-class heterogeneity.** Figure 3 Panel (a) indicates that the NME-demand elasticity varies across the eight major drug classes. We want to be sure that the shifts in TFP we identify are not driven by our restriction of homogeneous (i) investment responses, and (ii) output elasticities across these drug classes. To do so, we estimate investment and production functions where we allow both the investment response to demand ($\gamma_{c(j)}$) and output elasticity ($\beta_{c(j)}$) to vary across the PhRMA drug classes c (the eight groupings of the ATC-3 classes).

$$\log(R_{jy}) = \gamma_0 + \gamma_{c(j)}\mathbb{E}[\text{Market}]_{jy} + \delta_j + \tau_{t(y)} + \varepsilon_{jy}. \tag{B1}$$

FIGURE B1

TFP AND RENTS ALLOWING FOR HETEROGENEITY [Color figure can be viewed at wileyonlinelibrary.com]



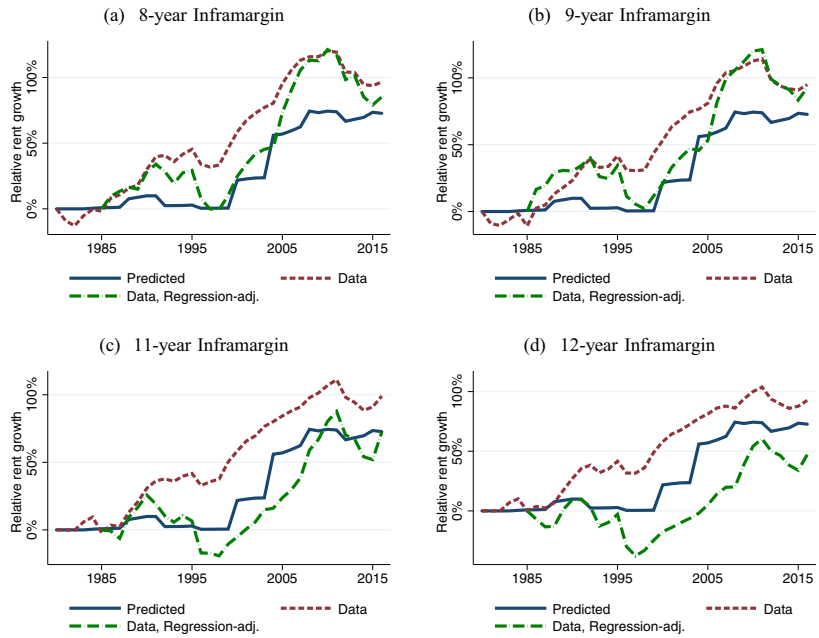
$$\text{NME}_{jy} = \frac{\exp\left(\alpha_{t(y)} + \beta_{c(j)} \log(R_{jy})\right)}{\sum_{v=1}^{Y} \exp\left(\alpha_{t(v)} + \beta_{c(j)} \log(R_{jv})\right)} + \epsilon_{jy}. \tag{B2}$$

Figure B1 Panel (a) reports the estimates of $\alpha_{t(y)}$ recovered from our 2SPS approach to estimating equations (B1) and (B2). Panel (b) recreates the Ricardian rent exercise with these new estimates. The results are very similar to the model that does not allow for this heterogeneity, with relative TFP declines near 100% and the predicted rents aligning very closely with our data-based proxies.

□ **Alternative rent trend specifications.**

FIGURE B2

RELATING DEMAND GROWTH TO PRODUCTIVITY AND RENTS [Color figure can be viewed at wileyonlinelibrary.com]



Note: Recreates Figure 7 Panel (c), comparing alternative windows of NME approvals for which rents are estimated in the data (for example, the 8-year inframargin plots data-based rent proxies using the count of NMEs that are within 8 years post- FDA approval). Thus, these numbers are varying the assumed effective market exclusivity time frame. Figure 7 is based on a 10-year window, and the predicted values are all equivalent, because they come from the regression model results.

References

Acemoglu, D., Finkelstein, A., and Notowidigdo, M.J. "Income and Health Spending: Evidence from Oil Price Shocks." *Review of Economics and Statistics*, vol. 95 (2013), pp. 1079–1095.

Acemoglu, D. and Linn, J. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics*, vol. 119 (2004), pp. 1049–1090.

Azoulay, P., Graff Zivin, J.S., Li, D., and Sampat, B.N. "Public R&D Investments and Private-Sector Patenting: Evidence from NIH Funding Rules." *Review of Economic Studies*, vol. 86 (2018), pp. 117–152.

Bhattacharya, J. and Packalen, M. "Opportunities and Benefits as Determinants of the Direction of Scientific Research." *Journal of Health Economics*, vol. 30 (2011), pp. 603–615.

Bloom, N., Jones, C.I., Van Reenen, J., and Webb, M. "Are Ideas Getting Harder to Find?" NBER Working Paper No. 23782, Department of Economics, Stanford University, 2018.

Blume-Kohout, M.E. and Sood, N. "Market Size and Innovation: Effects of Medicare Part D on Pharmaceutical Research and Development." *Journal of Public Economics*, vol. 97 (2013), pp. 327–336.

Blundell, R., Griffith, R., and Windmeijer, F. "Individual Effects and Dynamics in Count Data Models." *Journal of Econometrics*, vol. 108 (2002), pp. 113–131.

Budish, E., Roin, B.N., and Williams, H. "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials." *American Economic Review*, vol. 105 (2015), pp. 2044–2085.

Butler, D. "Translational Research: Crossing the Valley of Death." *Nature News*, vol. 453 (2008), pp. 840–842.

Collins, F.S. "Reengineering Translational Science: The Time Is Right." *Science Translational Medicine*, vol. 3 (2011), p. 90cm17.

Deloitte, L. "Measuring the Return from Pharmaceutical Innovation 2015." Tech. rep., Deloitte, LLP, 2015.

DiMasi, J.A., Grabowski, H.G., and Hansen, R.W. "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs." *Journal of Health Economics*, vol. 47 (2016), pp. 20–33.

DiMasi, J.A., Hansen, R.W., and Grabowski, H.G. "The Price of Innovation: New Estimates of Drug Development Costs." *Journal of Health Economics*, vol. 22 (2003), pp. 151–185.

DiMasi, J.A., Hansen, R.W., Grabowski, H.G., and Lasagna, L. "Cost of Innovation in the Pharmaceutical Industry." *Journal of Health Economics*, vol. 10 (1991), pp. 107–142.

Dubois, P., de Mouzon, O., Scott-Morton, F., and Seabright, P. "Market Size and Pharmaceutical Innovation." *RAND Journal of Economics*, vol. 46 (2015), pp. 844–871.

Duggan, M. and Morton, F.S. "The Effect of Medicare Part D on Pharmaceutical Prices and Utilization." *American Economic Review*, vol. 100 (2010), pp. 590–607.

Finkelstein, A. "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry." *Quarterly Journal of Economics*, vol. 119 (2004), pp. 527–564.

Getz, K.A. and Campo, R.A. "Trial Watch: Trends in Clinical Trial Design Complexity." *Nature Reviews Drug Discovery*, vol. 16 (2017), pp. 307–308.

Gordon, R.J. "Is US Economic Growth Over? Faltering Innovation Confronts the Six Headwinds." Working Paperw no. 18315, National Bureau of Economic Research, 2012.

Hall, R.E. and Jones, C.I. "The Value of Life and the Rise in Health Spending." *Quarterly Journal of Economics*, vol. 122 (2007), pp. 39–72.

Hegde, D. and Sampat, B. "Can Private Money Buy Public Science? Disease Group Lobbying and Federal Funding for Biomedical Research." *Management Science*, vol. 61 (2015), pp. 2281–2298.

Hewitt, J., Campbell, J.D., and Cacciotti, J. "Beyond the Shadow of a Drought." Tech. rep., Oliver Wyman, 2011.

Jones, B.F. "The Burden of Knowledge and the 'Death of the Renaissance man': Is Innovation Getting Harder?" *Review of Economic Studies*, vol. 76 (2009), pp. 283–317.

Jones, C.I. "R&D-Based Models of Economic Growth." *Journal of Political Economy*, vol. 103 (1995), pp. 759–784.

———. "The Shape of Production Functions and the Direction of Technical Change." *Quarterly Journal of Economics*, vol. 120 (2005), pp. 517–549.

Pammolli, F., Magazzini, L., and Riccaboni, M. "The Productivity Crisis in Pharmaceutical R&D." *Nature Reviews Drug Discovery*, vol. 10 (2011), pp. 428–438.

Restuccia, D. and Rogerson, R. "Misallocation and Productivity." *Review of Economic Dynamics*, vol. 1 (2013), pp. 1–10.

Ricardo, D. *On the Principles of Political Economy and Taxation*. J. Murray, 1817, London.

Ruffolo, R. "Why Has R&D Productivity Declined in the Pharmaceutical Industry?" *Expert Opinion on Drug Discovery*, vol. 1 (2006), pp. 99–102.

Scannell, J.W., Blanckley, A., Boldon, H., and Warrington, B. "Diagnosing the Decline in Pharmaceutical R&D Efficiency." *Nature Reviews Drug Discovery*, vol. 11 (2012), pp. 191–200.

Schmookler, J. *Invention and Economic Growth*. Cambridge, MA: Harvard University Press, 1966.

Terza, J.V., Basu, A., and Rathouz, P.J. "Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling." *Journal of Health Economics*, vol. 27 (2008), pp. 531–543.

Toole, A.A. "The Impact of Public Basic Research on Industrial Innovation: Evidence from the Pharmaceutical Industry." *Research Policy*, vol. 41 (2012), pp. 1–12.

Vos, T., Barber, R.M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., Charlson, F., Davis, A., Degenhardt, L., Dicker, D., et al. "Global, Regional, and National Incidence, Prevalence, and Years Lived with Disability for 301 Acute and Chronic Diseases and Injuries in 188 Countries, 1990–2013: A Systematic Analysis for the Global Burden of Disease Study 2013." *Lancet*, vol. 386 (2015), pp. 743–800.

Ward, D.J., Hammond, E., Linden-Phillips, L., and Stevens, A.J. "Trends in Clinical Development Timeframes for Antiviral Drugs Launched in the UK, 1981–2014: A Retrospective Observational Study." *BMJ Open*, vol. 5 (2015), p. e009333.