Health Economics Series
No. 2017-06

Endogenous Productivity of Demand-Induced R&D: Evidence from Pharmaceuticals

Mark Pauly and Kyle Myers

September 2017

JEL Codes: D20, L10, L65, O31

Becker Friedman Institute
for Research in Economics

Contact:
773.702.5599
bfi@uchicago.edu
bfi.uchicago.edu

Endogenous Productivity of Demand-Induced R&D: Evidence from Pharmaceuticals

Mark Pauly & Kyle Myers*

September 24, 2017

Abstract

This paper examines trends in the productivity of the pharmaceutical sector over the past three decades. Motivated by Ricardo's insight regarding demand-driven productivity in settings of scarce resources, we examine the industry's aggregate R&D production function. Using exogenous demand shocks to instrument investments, we find that demand growth explains roughly half of R&D growth and amongst this demand-induced R&D, the industry's returns to scale have been very stable while total factor productivity has declined significantly. Suggestive evidence based on these estimates is in line with Ricardo's prediction that productivity and rents are endogenous to demand.

Keywords: innovation; productivity; pharmaceuticals
JEL Codes: D20, L10, L65, O31

*Pauly: Health Care Management Department, The Wharton School, University of Pennsylvania. 208 Colonial Penn Center, 3641 Locust Walk, Philadelphia, PA 19104; pauly@wharton.upenn.edu. Myers (Corresponding Author): Health Care Management Department, The Wharton School, University of Pennsylvania. G7 Colonial Penn Center, 3641 Locust Walk, Philadelphia, PA 19104; myersky@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, Neeraj Sood and other attendees at the 2016 Bates White Life Science Symposium. All errors are our own.

1 Introduction

Considerable criticism has been expressed about the growing research and development costs for new drugs. These critiques often cite estimates of the average out-of-pocket costs to bring a new drug to market, which have increased in real terms from $230M in the 1980s to $540M in the 1990s to $1,300M in the 2000s (DiMasi et al. 1991, 2003, 2016; $-2012)[1]. More broadly, the declining pace of innovation and growing costs of R&D per output have been documented across the U.S. economy (Jones 1995; Gordon 2012; Bloom et al. 2017). Still, it remains unclear to what extent these trends are driven by forces such as mismanagement, regulatory burdens, or are instead the expected outcomes of rational firms making investments in more "difficult", but still highly demanded ideas[2].

Instead of pinpointing specifics of these declines, the goal of this paper is to reframe the discussion to one of demand-induced investment amidst scarce ideas. We do this by connecting the classic notion 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 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

[1] See, for example, Hewitt et al. (2011), or Scannell et al. (2012). These R&D figures are frequently cited in public discussions of pharmaceutical prices, although rarely following standard economic logic - ex-post optimal prices are independent of ex-ante sunk costs.

[2] For instance, when discussing semiconductors and growing costs of sustaining Moore's Law, Bloom et al. (2017) 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 TFP there in order to achieve the gains associated with Moore's Law".

demographics-driven measure of demand to instrument . We emphasize how policy recommendations depend on whether the observed declines in productivity are driven by changes in α or β. For instance, the elasticity of new drugs (i.e. $\%\Delta N$) with respect to market size - one measure of the industry's ability to "meet demand" - depends only on β.

The pharmaceutical sector provides both an inherently important and empirically ideal setting to study R&D dynamics since we can (1) identify exogenous demand shocks, (2) connect these shocks to R&D investments and new products, and (3) separately identify changes to TFP and output elasticities. To do this, we utilize therapeutic-class-specific data on U.S. consumer drug expenditures, private U.S.-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 - suggesting that first-order supply-side frictions on marginal R&D investments (declines in β) are unlikely.

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 total investments to illustrate the role of potential demand in stimulating private R&D. Notably, demand alone can explain roughly half of the growth in total R&D since 1980. This also makes clear the scope of our two-stage analyses - we can only explore changes in the productivity of R&D dollars invested in response to demand. Still, our results are relevant for any policies that influence demand-side

[3]This has some reasonable implications, e.g. 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.

features of this market (e.g. insurance, drug vouchers, market exclusivity rules).

Using these predicted values, we estimate a version of Eq. 1 over the full sample. In summary, our estimates fail to reject the null that output elasticity (β) is stable - the industry's returns to scale haven't significantly changed. However, we do identify significant declines in TFP (α) on the order of 200%. In terms of demand responsiveness, these estimates imply that a growing market (per drug class) in 2012 will still induce the same, proportional increase in new drugs as in 1980. But compared to 1980's averages, all 2012 markets received roughly one quarter 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). We fail to find any significant relationships related to our main analyses, which is a part of our story - because we don't find evidence that the NIH's supply of research (as one source of the industry's TFP) is endogenous to demand while 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 in line with his two major predictions: during periods of demand growth (1) firms, facing scarce inputs, invest in ideas that are less "fertile"; and (2) earn larger rents as a function of these declines in fertility.

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[4].

[4]Finkelstein (2004) explores a policy change related to the value of vaccines and finds a 1% increase

A large body of literature has also investigated the supply-side forces of innovation such as technological opportunities, spillovers and competition (e.g. Scherer 1965; Jaffe 1988; Bloom et al. 2017). But to reiterate, we focus here on demand-driven R&D investments under the assumption that the demand shocks we observe are orthogonal to any supply-side changes (e.g. new drug ideas created by basic science). In Section 5, we examine research grant funding at the NIH and find no evidence to suggest this orthogonality assumption is invalid.

Our analysis also complements recent examinations of the productivity slowdown. In closely related work, Bloom et al. (2017) 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 Eq. (1), they find declining R&D productivity within many sectors including semiconductors, agriculture, and (using similar data as us) pharmaceuticals. This paper is a useful joinder to theirs in that: (1) Bloom et al. (2017) cannot disentangle whether these declines are due to parameters or while we can, and (2) our consideration for the endogeneity of scarce R&D inputs with respect to demand provides some insight as to one likely cause of the productivity declines - an explanation they hint at.

The remainder of the paper 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.

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.

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. But 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. 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 will consider the implications of this proposition later[5].

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 potentially profitable. But 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.

[5]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.

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 Eq. 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 (e.g. a Hicks-neutral production shift), while output elasticities are held fixed. However, in the context of R&D, such an assumption warrants testing since the totality of ideas is (obviously) unknown. Thus, we remain agnostic as to how a scarcity of ideas would influence our 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 et al. (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 clinical trial costs[6]. Looking across time, increases in the length and complexity of clinical trials, which would decrease the value of all ideas, has been observed over the past decades. Regulatory burdens such as longer approval times and larger post-marketing surveillance efforts have also been cited as sources of larger costs in the industry (Kola and Landis 2004). 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.

[6]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 et al. 2015).

3 Empirical approach

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

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

where years are grouped into 4-year time periods t and R_{jt} is prior private R&D investments as described below[7]. This grouping of years into time periods facilitates our estimation of the parameters given the relatively small number of observations within each year, and follows AL04's original analyses. The Poisson specification accommodates the count nature of our dependent variable and focuses on the rate of innovation. 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[8]. Note also that Eq. 5.5 does not directly identify TFP levels as written in Eq. 1, but instead the α_t parameters estimate 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 Eq. 5.5 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 demand in (then future) year y per the

[7] 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.

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

investment function

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

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 Eq. 3 which in turn determines NME output per Eq. 5.5. Our motivation and process for lag (l) selection is described in Section 4.2 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 as outlined by Blundell et al. (2002)[9]. 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, this method of two-stage-predictor-substitution (2SPS) in nonlinear models has been shown to potentially suffer from bias (Terza et al. 2008). Ideally, we would jointly estimate the equations as a single non-linear 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 Eqs. 5.5 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

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

implies that managers cannot predict any productivity declines or improvements[10]. Given the multi-year lags between project initiation and eventual success or failure, this implication is reasonable.

Therefore, any changes in α and/or β that we identify are only with respect to the productivity of the private R&D instrumented by our measure of demand. Still, the variation we can explore is important because it speaks directly to R&D dynamics influenced by market size - a variable often targeted by health and innovation policies whether it be insurance coverage or patent rules[11]. Although this is a study of pharmaceutical productivity, we are examining variation likely the most generalizable to other industries.

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 demand. Table 1 summarizes the three key variables.

4.1 Drug classes & FDA approvals

Determining the specificity of drug classes for our empirical analysis is a tradeoff between the ability to control for important fixed differences across classes (e.g. underlying scientific potential, stable 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 since drug-level data is available. As a compromise, we use the Anatomical Ther-

[10] If anything, it is plausible there were unrealistically optimistic expectations of success. In fact, per Figure 1 R&D investments grew at a rate greater than expected given firms' estimated responsiveness to demand.

[11] Insurance coverage generally shields consumers from real prices and patents promote monopoly pricing, both affecting potential market size in terms of revenues.

apeutic Chemical (ATC) Classification System as a guide, we matched the eight industry R&D categories to the corresponding 65 subgroups of the ATC hierarchy as shown in Appendix 8.1.

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 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.

The count of NME approvals from 1987 to 2013 was constructed from the Drugs@FDA database. The database does provide information as to the sponsoring firm of each drug approval; however, it is prohibitive to track 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 ATC codes.

4.2 Private R&D expenditures

Estimates of industry-wide investments in pharmaceutical R&D from 1970 to 2013 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[12], 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 while 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. Additionally, because PhRMA underrepresents the industry as a whole we will likely underestimate real industry-wide investments. So long as any level of misrepresentation is fixed over time, it will not bias our estimates of marginal productivity.

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 1 and 12 years (given average reported development times of 10-12 years over this sample) and choose the lag from the model with the largest F statistic. Appendix 8.2 reports the results of these regressions, from which we choose a seven-year lag.

All private investments are deflated using the Biomedical Research and Development Price Index (BRDPI) since effective productivity of the industry's inputs and not nominal investments.

[12]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.

4.3 Exogenous potential demand

In their initial analyses of market size and innovation in the pharmaceutical sector, AL04 develop a plausibly exogenous measure market size they term "potential demand". 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 potential demand for drug category j in year y is given by

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

where a is a set of 5-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 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. Unweighted estimates are also reported.

AL04 provide evidence that within-group drug expenditure shares are relatively constant over time while the country's demographics are not. Intuitively, illnesses and the medications used to treat them often effect 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 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 2013, therefore we extend the average

calculated during this timeframe to our full timeframe of 1985 to 2013. CPS data is available for the full length of our study with all data deflated using the Consumer Price Index.

5 Main productivity results

5.1 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 Eq. 5.5 but using the exogenous demand measure as the dependent variable. Table 2 presents the results first omitting the time-period controls (Col. 1), then including the time controls (Col. 2) and allowing the coefficient on market size to vary pre-post 2000 (Col. 3). When including time controls, our point estimates are well within the bounds of AL04's estimates[13]. Column 3 suggests that the NME-demand elasticity did not change after 2000, while the average rate of NMEs across all classes declined substantially as evidenced by the significant drops in the time-specific intercepts. This gives initial evidence that in terms of Eq. 1, any productivity changes were likely concentrated within the TFP term (α).

Figure 2 plots estimates of the NME-demand elasticity when allowed to vary over time period (Panel A) and across the 8 major R&D classes (Panel B). Panel A provides further evidence that this elasticity did not substantially differ from 1980 to 2013 and suggests that declines in the industry's output elasticity (β) are unlikely. Panel B indicates that there is significant heterogeneity in the NME-demand elasticity across therapeutic classes. From these results alone we cannot separate the extent to which this is due to productivity variation across classes (β varies by j) or differences with respect to the degree of R&D inducement (γ_D varies

[13]See AL04 Table 2, Panel C, Column 1 where they report a NME-demand elasticity of 3.54 (S.E. = 1.19).

by j). We explore this further in Section 5.4 and also take steps in the following regressions to accommodate for this cross-sectional source of variation by allowing parameters to vary across the R&D classes.

5.2 First stage: R&D investments and market size

Table 2, Columns (1) and (2) presents the "first stage" results where we regress prior private R&D on current values of the demand measure per Eq. (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 a 12% increase in R&D spending is made in advance of a 10% exogenous growth in market size. Column 2 presents a similar pattern to that found in Figure 2, Panel A where there is not significant within-panel variation in this elasticity (albeit over a shorter timeframe). 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 indicating that roughly half of the growth in annual R&D spending since 1980 is due to demand growth. We also note that one feature of demand in this sector that the exogenous market size measure does not accommodate is an income elasticity of health above 1. In order maintain exogeneity, the measure implies that a constant share of income is spent on health. Although empirical evidence is mixed (Acemoglu et al. 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 under predicting real demand growth.

5.3 Two Stage: R&D productivity

Before investigating productivity over the full sample, we first must validate our 2SPS approach within the timeframe that we do not face the main data limitation. 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 (Cols. 3-4), GMM Poisson where we estimate Eqs. (3) and (4) together (Cols. 5-6) and 2SPS Poisson where we use predicted R&D investments from the endogenous investment Eq. (4) to estimate the production function Eq. (3) (Cols. 7-8). When using the exogenous demand measure to instrument R&D spending, we identify a ranging from 2.1 to 2.9, which imply an average cost per marginal NME of roughly $400M to $800M. As an indication of accuracy, these magnitudes are very similar to the accounting-based cost estimates over similar time periods, which range from $200M to $600M (DiMasi et al. 1991, 2003).

But to reiterate, our focus is the evolution of these costs over time. And since 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 get a sense of the magnitude of this bias and more importantly whether this magnitude changes over time, we can compare Columns (5-6) to Columns (7-8) in Table 2. On average the 2SPS estimates are about 10-30% larger than the GMM estimates; however, comparing Columns (6) and (8) reveals that both specifications produce very similar estimates over time with each estimating an insignificant annual 0-1% declines in β. Thus, 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 results are presented in Figure 3. Panels A and B present the results of estimating the 2SPS version of the investment and production functions and allowing the and coefficients, respectively, to vary over time periods. Virtually all of the changes in productivity

arise within the (TFP) coefficient while the returns to scale (output elasticity) of this industry are stable. This pattern is in line with the results presented in Figure 3 Panel A where the NME-demand elasticity (which depends only on) is also very stable[14].

5.4 Heterogeneity & robustness

Motivated by the heterogeneity in the NME-demand elasticity across R&D classes (Figure 2, Panel B), we also estimated 2SPS models where we allowed both the demand-investment relationship (γ_D) and output elasticity (β) to vary by major R&D class. Figure 4 presents these results and give some insight as to potential causes of the variation in NME-demand elasticity. For example, the two classes with the lowest NME-demand elasticities (dermatological and gastrointestinal) - are both estimated to have the largest output elasticities but the smallest investment elasticities. This overall pattern of NME-demand elasticities being more correlated with investment elasticities as opposed to output elasticities suggests that there are other significant market features - conditional on raw market size - that are influencing firms' investment decisions. Whether such features are driven by differences in consumer-level features or competitive forces we leave to future research.

Table 4 presents a number of robustness tests based on estimating the 2SPS model over the full sample where β is allowed to vary over time while α is not. The pattern of declines in is persistent across the variety of specifications changing the lag structure of private investments or aggregating up the hierarchy of drug classes (originally 65) to match the eight major R&D classes.

[14]Note, however, that since we are investigated the flow rate of new drugs, a fixed does not necessarily imply a fixed marginal cost per NME. At the sample means, the estimates presented in Figure 3 Panel A suggest a marginal cost per NME of roughly $589M in 1980, $528M in 1990, $622M in 2000 and $1,041M in 2010. Again, these estimates are in line with those from the DiMasi et al. (1991, 2003, 2016) series of accounting-based figures.

5.5 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 rate of new drug approvals outside of private R&D. The most obvious concern is that other funders of biomedical research, especially research that may be used as an input to private R&D, 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.

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). Perhaps most 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 on patents marketable drug candidates (Azoulay et al. 2016). While Azoulay et al.'s (2016) 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. (2016), 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 (A) correlations

between NIH funding and the demand shocks that would suggest an endogenous public response, (B) correlations between NIH funding and the demand-induced private R&D investments that would also suggest endogeneity, and (C) an additive production function to recover and compare the marginal productivity of public and private R&D[15].

Data on NIH extramural grant awards from 1965 to 2013 was constructed based on data available from the NIH ExPORTER data files as well as a Freedom of Information Act request for 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 employ a keyword based approach to categorizing grants to the 16 major ATC classes as outlined in Appendix 8.1. As with private R&D, all NIH investments are deflated using the BRDPI to ensure comparison of effective inputs.

Table 5 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, 16 and 18 years prior to the focal year of NME approvals. Across this range of lags, Panels A and B reveal no significant correlation between either the demand shocks or the private investments they induce. This provides evidence that the induced private R&D investments we observe are due solely to the exogenous demand shocks.

Panel C of Table 5 estimates a version of the NME production function (Eq. 5.5) now including NIH funding. The relationship between NIH funding and NME output appears somewhat sensitive to the lag specification, with coefficients ranging from 0.25 to 0.49 (i.e.

[15]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.

a 10% increase in NIH funding in year $y - l$ is correlated with a 4.9% increase in NME output in year y). The coefficient on predicted private R&D is consistent with estimates not including NIH funding (2.1 to 2.5), which is 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. (2016)). Based on the average coefficients for each source of investments, the estimates suggest that one additional NME could be produced with $720 million in private R&D or $3.09 billion in NIH funding.

We caution interpretation of these 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 chief source of friction.

6 Discussion

6.1 Revisiting Ricardo's predictions

To summarize our findings at this point, we identify significant declines in TFP for the industry since 1980, while 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

might explain patterns of this industry. The first straightforward prediction of the model is that to the extent 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.

The second prediction of Ricardo's theory is that these demand-driven productivity declines will lead to corresponding increases in the rents accrued to firms, as a function of the extent of the productivity declines. To clarify, first note the fact that TFP is inversely proportional to the slope of the supply curve - a more productive industry requires less demand growth to increase output. Second, note that following a demand shock, the change in rents is proportional to the slope of the supply curve - if supply increases slower than demand as it would in a world of scarce inputs, the difference in growth rates will be reflected in rents. Together, this implies that rent growth is proportional to the product of demand growth and the inverse of productivity growth, which we use to predict rent growth based on our estimates of demand and TFP growth: $\%\Delta \text{Rent}_y^1 = \%\Delta \mathbb{E}[\text{Demand}]_y \times \%\Delta \text{TFP}_y$.

Ideally, we would identify many market-time-specific estimates of demand, productivity and rents. Then we could systematically relate demand shocks to changes in productivity examine the connection to rents. In our case (with only eight estimates of TFP changes) this is not possible, and in general would necessitate an unusually large and detailed dataset. But 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. Still, we present this discussion only as suggestive and caution against making any generalizations beyond this specific industry or this specific time period.

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 point. 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 U.S. over this time period.

Estimating rents in the Ricardian sense is less straightforward. Ideally, a measure of these rents would relate time-varying drug-specific prices to the original average R&D costs required to bring those same drugs to market. Since we can only observe total annual NME approvals and sales, we construct a data-based rent proxy per

$$\%\Delta \text{Rent}_y^2 \equiv \%\Delta \left[\frac{\text{Revenue}_y}{\mathbb{E}[\text{Demand}]_y} / \sum_{y'=y-10}^{y} \text{NME}_{y'} \right] \tag{5}$$

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 (the average effective patent life) ensures two things: (1) that we are comparing drugs developed during similar timeframes and likely have similar development costs, and (2) account for the number of new patent-protected products on the market at a given time. 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.

Figure 5 plots our estimates of demand, TFP and rents over time. In line with the first prediction, the nonlinear pattern of demand growth since 1980 (Panel A) is reflected as an inverse pattern of TFP declines (Panel B). For instance, the pauses in demand growth of the early 1990's and most recent years are reflected by relatively stable periods of TFP at both times. Panel C of Figure 5 plots the proxy estimates of rent growth per Eq. 5 alongside our predicted rents per Ricardo. Both in terms of absolute magnitudes and relative changes, the estimated and predicted growth in industry rents track very closely. Overall, these trends

are in line with what Ricardo would expect if new drug ideas - as the source of TFP - are indeed scarce. But to reiterate, we note these patterns only in terms of preliminary evidence that does not clearly reject our null hypotheses of Ricardo's model and hope future research may further investigate these relationships.

The suggestive result that a majority of the changes in rents accrued to this industry can be explained by this theory of scarce inputs has important implications for policies focused on drug prices. Traditionally, policymakers intent on lowering drug prices have discussed mechanisms such as implementing cost-effectiveness thresholds (at least, outside the U.S.) 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.

Given the aggregate nature of our data and analyses, we cannot speak directly to what exactly "TFP" is outside of its empirical representation in Eq. 1. But 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 public and/or private investments in basic science. Recall Section 5.5 and the lack of any relationship between NIH investments and future market size; it is then not surprising then that we find evidence of ideas being in increasingly short supply as markets expand.

6.2 Conclusion

Utilizing demographics-driven changes in demand to instrument R&D investments, this paper estimates the productivity of the pharmaceutical sector over the past three decades. Not only is this demand instrument useful for its econometric properties, but because the local productivity estimates identified are based on investments made directly in response to changes in market size - a lever that policymakers often pull in hopes of stimulating R&D.

But 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 while output elasticity has remained very stable, TFP has declined by roughly 200%. In line with Ricardo's predictions, the periods of greatest TFP declines coincided with the greatest demand growth, which together led to increased rents. 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.

References

Acemoglu, D., Finkelstein, A., and Notowidigdo, M. J. (2013). Income and health spending: Evidence from oil price shocks. *Review of Economics and Statistics*, 95(4):1079–1095.

Acemoglu, D. and Linn, J. (2004). Market size in innovation: Theory and evidence from the pharmaceutical industry. *The Quarterly journal of economics*, 119(3):1049–1090.

Azoulay, P., Graff-Zivin, J., Li, D., and Sampat, B. (2016). Public r&d investments and private sector patenting: Evidence from nih funding rules. *Mimeo.*

Bhattacharya, J. and Packalen, M. (2011). Opportunities and benefits as determinants of the direction of scientific research. *Journal of health economics*, 30(4):603–615.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2017). Are ideas getting harder to find? *Unpublished paper, Stanford University.*

Blume-Kohout, M. E. and Sood, N. (2013). Market size and innovation: Effects of medicare part d on pharmaceutical research and development. *Journal of public economics*, 97:327–336.

Blundell, R., Griffith, R., and Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics*, 108(1):113–131.

Budish, E., Roin, B. N., and Williams, H. (2015). Do firms underinvest in long-term research? evidence from cancer clinical trials. *The American economic review*, 105(7):2044–2085.

DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics*, 47:20–33.

DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of health economics*, 22(2):151–185.

DiMasi, J. A., Hansen, R. W., Grabowski, H. G., and Lasagna, L. (1991). Cost of innovation in the pharmaceutical industry. *Journal of health economics*, 10(2):107–142.

Dubois, P., de Mouzon, O., Scott-Morton, F., and Seabright, P. (2015). Market size and pharmaceutical innovation. *The RAND Journal of Economics*, 46(4):844–871.

Duggan, M. and Morton, F. S. (2010). The effect of medicare part d on pharmaceutical prices and utilization. *The American Economic Review*, 100(1):590–607.

Finkelstein, A. (2004). Static and dynamic effects of health policy: Evidence from the vaccine industry. *The Quarterly Journal of Economics*, 119(2):527–564.

Gordon, R. J. (2012). Is us economic growth over? faltering innovation confronts the six headwinds. Technical report, National Bureau of Economic Research.

Hall, R. E. and Jones, C. I. (2007). The value of life and the rise in health spending. *Technology*, 39.

Hegde, D. and Sampat, B. (2015). Can private money buy public science? disease group lobbying and federal funding for biomedical research. *Management Science*, 61(10):2281–2298.

Hewitt, J., Campbell, J. D., and Cacciotti, J. (2011). Beyond the shadow of a drought. *Oliver Wyman*.

Jaffe, A. B. (1988). Demand and supply influences in r & d intensity and productivity growth. *The Review of Economics and Statistics*, pages 431–437.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies*, 76(1):283–317.

Jones, C. I. (1995). R&d-based models of economic growth. *Journal of political Economy*, 103(4):759–784.

Jones, C. I. (2005). The shape of production functions and the direction of technical change. *The Quarterly Journal of Economics*, 120(2):517–549.

Kola, I. and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery*, 3(8):711–716.

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

Scannell, J. W., Blanckley, A., Boldon, H., and Warrington, B. (2012). Diagnosing the decline in pharmaceutical r&d efficiency. *Nature reviews Drug discovery*, 11(3):191–200.

Scherer, F. M. (1965). Firm size, market structure, opportunity, and the output of patented inventions. *The American Economic Review*, 55(5):1097–1125.

Schmookler, J. (1966). Invention and economic growth.

Terza, J. V., Basu, A., and Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, 27(3):531–543.

Toole, A. A. (2012). The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry. *Research Policy*, 41(1):1–12.

Tables

Table 1: Summary Statistics

	(1) Full Sample (1980-2013)	(2) 1980's	(3) 1990's	(4) 2000's
NME_{jy}	0.359	0.258	0.446	0.369
	(0.777)	(0.572)	(0.826)	(0.857)
Private $\text{R\&D}_{j(y-7)}$ ($B)	0.956	0.517	1.173	1.854
	(0.772)	(0.282)	(0.804)	(1.056)
$\mathbb{E}[\text{Demand}]_{jy}$ ($B)	2.194	1.519	2.019	2.801
	(3.292)	(2.168)	(2.891)	(4.038)
Obs. (Class-Years)	2210	650	650	910

Note: Mean values within class j year y, s.d. in parentheses. R&D and Demand values deflated by BRDPI and CPI, respectively.

Table 2: Reduced form NME production function

	NME_{jy}		
	(1)	(2)	(3)
$\mathbb{E}[\text{Demand}]_{jy}$	0.453	2.805***	
	(0.304)	(0.912)	
$\mathbb{E}[\text{Demand}]_{jy}$ × Pre-2000			3.025***
			(0.941)
$\mathbb{E}[\text{Demand}]_{jy}$ × Post-2000			3.013***
			(0.938)
Year_{84-87}		-0.181	-0.201
		(0.184)	(0.184)
Year_{88-91}		-0.607**	-0.659**
		(0.282)	(0.284)
Year_{92-95}		-0.488	-0.560
		(0.355)	(0.354)
Year_{96-99}		-0.545	-0.639
		(0.462)	(0.461)
Year_{00-03}		-1.101*	-1.167*
		(0.648)	(0.644)
Year_{04-07}		-1.662**	-1.548**
		(0.682)	(0.681)
Year_{08-11}		-1.907**	-1.809**
		(0.757)	(0.756)
Year_{12-13}		-1.826**	-1.738**
		(0.828)	(0.828)
Obs.	2210	2210	2210

Note: Standard errors in parentheses, clustered at R&D-class level. Estimated as a conditional (per drug class j) Poisson model. "Year" variables are time period indicators for the range denoted in subscript. * p<0.10, ** p<0.05, *** p<0.01

Table 3: Main investment and productivity results

	Private R&D$_{j(y-7)}$		NME$_{jy}$					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{E}[\text{Demand}]_{jy}$	1.200***							
	(0.159)							
$\mathbb{E}[\text{Demand}]_{jy} \times$ Year$_{(80-83)}$		1.174***						
		(0.146)						
$\mathbb{E}[\text{Demand}]_{jy} \times$ Year$_{(84-87)}$		1.179***						
		(0.143)						
$\mathbb{E}[\text{Demand}]_{jy} \times$ Year$_{(88-91)}$		1.195***						
		(0.144)						
$\mathbb{E}[\text{Demand}]_{jy} \times$ Year$_{(92-95)}$		1.221***						
		(0.144)						
$\mathbb{E}[\text{Demand}]_{jy} \times$ Year$_{(96-00)}$		1.223***						
		(0.151)						
Private R&D$_{j(y-7)}$			0.788***		2.558***		2.852***	
			(0.258)		(0.843)		(0.823)	
Private R&D$_{j(y-7)} \times$ Year$_{(80-83)}$				0.808***		2.201***		2.913***
				(0.271)		(0.822)		(0.820)
Private R&D$_{j(y-7)} \times$ Year$_{(84-87)}$				0.802***		2.197***		2.887***
				(0.269)		(0.814)		(0.812)
Private R&D$_{j(y-7)} \times$ Year$_{(88-91)}$				0.805***		2.193***		2.890***
				(0.263)		(0.797)		(0.808)
Private R&D$_{j(y-7)} \times$ Year$_{(92-95)}$				0.800***		2.157***		2.884***
				(0.258)		(0.785)		(0.803)
Private R&D$_{j(y-7)} \times$ Year$_{(96-00)}$				0.786***		2.110***		2.854***
				(0.254)		(0.771)		(0.793)
Implied Marginal $\frac{\$M}{NME}$			\$2,235		\$798		\$474	
Obs.	1170	1170	1080	1080	1170	1170	1080	1080
Time F.E.	Y	Y	Y	Y	Y	Y	Y	Y
I.V.					Y	Y	Y	Y
Spec.	OLS	OLS	Poisson	Poisson	GMM	GMM	2SPS Poisson	2SPS Poisson

Note: Standard errors in parentheses, clustered at R&D class level. Columns (1) and (2) provide first-stage estimates of Eq. 3. Columns (3-8) provide estimates of Eq. 5.5. * p<0.10, ** p<0.05, *** p<0.01. Implied Marginal $\frac{\$M}{NME}$ denotes the average of the marginal costs per NME based on the coefficient estimates in the respective columns that use the same specifications.

Table 4: Robustness tests: Productivity results

	NME_{jy}				
	(1)	(2)	(3)	(4)	(5)
Pred. Private $R\&D_{j(y-l)}$	2.320***	1.710***	1.489***	1.927***	1.589***
	(0.787)	(0.556)	(0.484)	(0.131)	(0.0566)
$Year_{(84-87)}$	-0.105	-0.181	-0.181	-0.114***	-0.0247
	(0.157)	(0.184)	(0.184)	(0.0214)	(0.0183)
$Year_{(88-91)}$	-0.518*	-0.607**	-0.607**	-0.528***	-0.333***
	(0.270)	(0.282)	(0.282)	(0.0442)	(0.0276)
$Year_{(92-95)}$	-0.397	-0.488	-0.488	-0.417***	-0.161***
	(0.373)	(0.355)	(0.355)	(0.0592)	(0.0364)
$Year_{(96-99)}$	-0.457	-0.545	-0.545	-0.543***	-0.124***
	(0.450)	(0.462)	(0.462)	(0.0736)	(0.0432)
$Year_{(00-03)}$	-1.003	-1.101*	-1.101*	-1.157***	-0.554***
	(0.661)	(0.648)	(0.648)	(0.109)	(0.0587)
$Year_{(04-07)}$	-1.535**	-1.662**	-1.662**	-1.678***	-1.034***
	(0.642)	(0.682)	(0.682)	(0.107)	(0.0612)
$Year_{(08-11)}$	-1.758**	-1.907**	-1.907**	-1.779***	-1.168***
	(0.721)	(0.757)	(0.757)	(0.119)	(0.0631)
$Year_{(12-13)}$	-1.682**	-1.826**	-1.826**	-1.944***	-1.100***
	(0.798)	(0.828)	(0.828)	(0.133)	(0.0798)
Obs.	2210	2210	2210	544	272
Weights		Y	Y	Y	Y
R&D lag (l)	7	10	12	7	7
Num. class (j)	65	65	65	16	8

Note: Standard errors in parentheses, clustered at R&D class level. * p<0.10, ** p<0.05, *** p<0.01. Weights are based on the number of observations used when constructing the demand instrument. The 16-class categorization scheme is based on the numerical indexed categories listed in the Appendix. The 8-class scheme is based on the eight major R&D classes listed in the PhRMA annual reports.

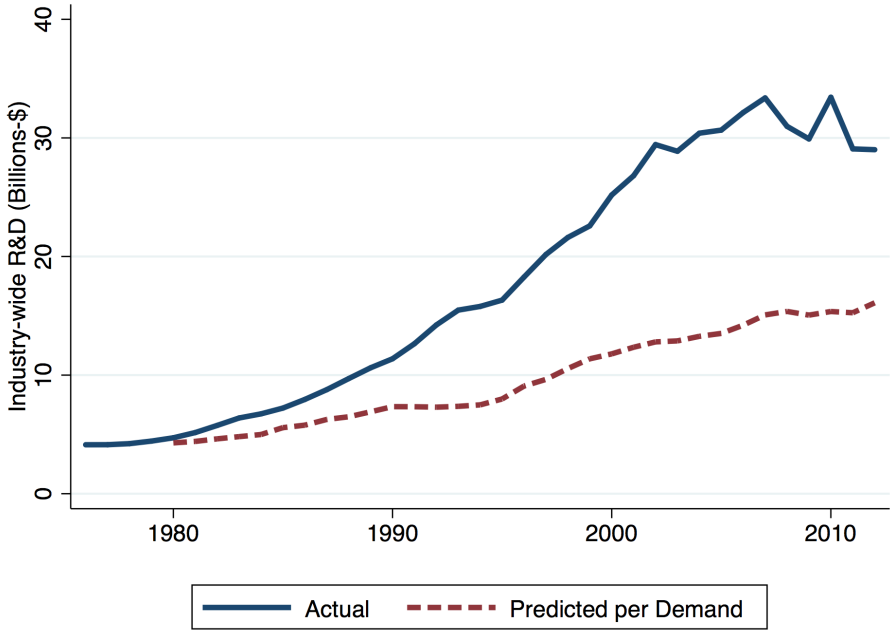
Table 5: Supply-side NIH Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Role of Demand							
Dep. Var. = $\text{NIH}_{j(y-l)}$; OLS							
$\mathbb{E}[\text{Demand}]_{jy}$	0.260	0.369*	0.187	0.373	-0.325	0.0651	-0.344
	(0.224)	(0.219)	(0.244)	(0.278)	(0.343)	(0.340)	(0.318)
Panel B: Role of Demand-Induced Private R&D							
Dep. Var. = $\text{NIH}_{j(y-l)}$; OLS							
Private $\widehat{\text{R\&D}}_{j(y-7)}$	0.217	0.308	0.156	0.311	-0.271	0.0543	-0.287
	(0.318)	(0.316)	(0.339)	(0.393)	(0.497)	(0.575)	(0.479)
Panel C: Private + Public NME Production Function							
Dep. Var. = NME_{jy}; Poisson							
Private $\widehat{\text{R\&D}}_{j(y-7)}$	2.137***	2.204***	2.250***	2.249***	2.489***	2.314***	2.538***
	(0.731)	(0.737)	(0.726)	(0.728)	(0.738)	(0.742)	(0.754)
$\text{NIH}_{j(y-l)}$	0.474**	0.434**	0.489**	0.438***	0.247	0.278**	0.333**
	(0.196)	(0.197)	(0.193)	(0.157)	(0.153)	(0.127)	(0.160)
Obs.	2210	2210	2210	2210	2210	2112	1984
R&D lag (l)	6	8	10	12	14	16	18

Note: Standard errors in parentheses, clustered at R&D class level. Private $\widehat{\text{R\&D}}_{j(y-7)}$ denotes the 2SPS estimates 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

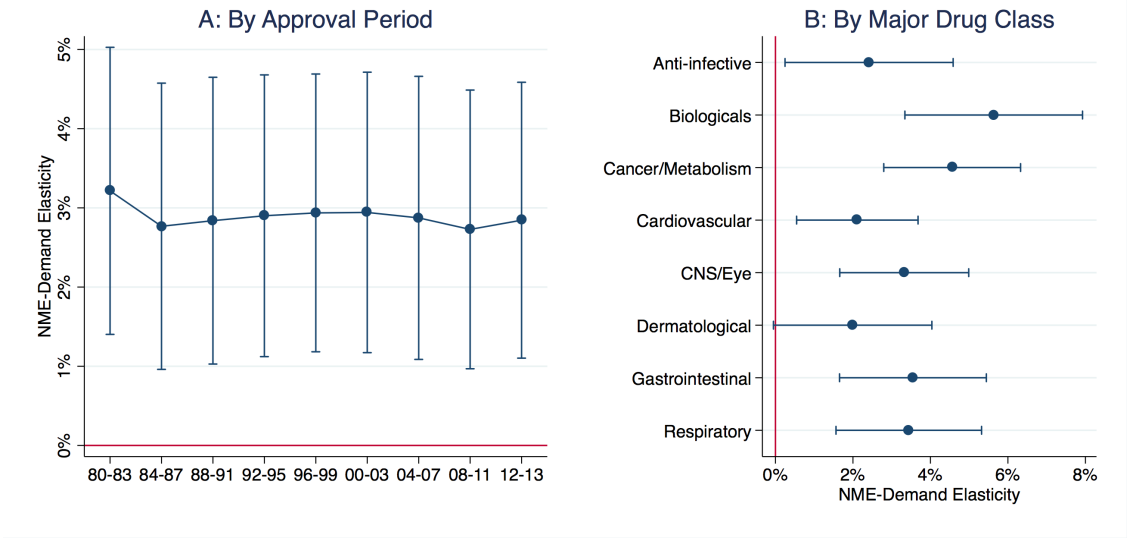
7 Figures

Figure 1: Actual and Predicted R&D Trends



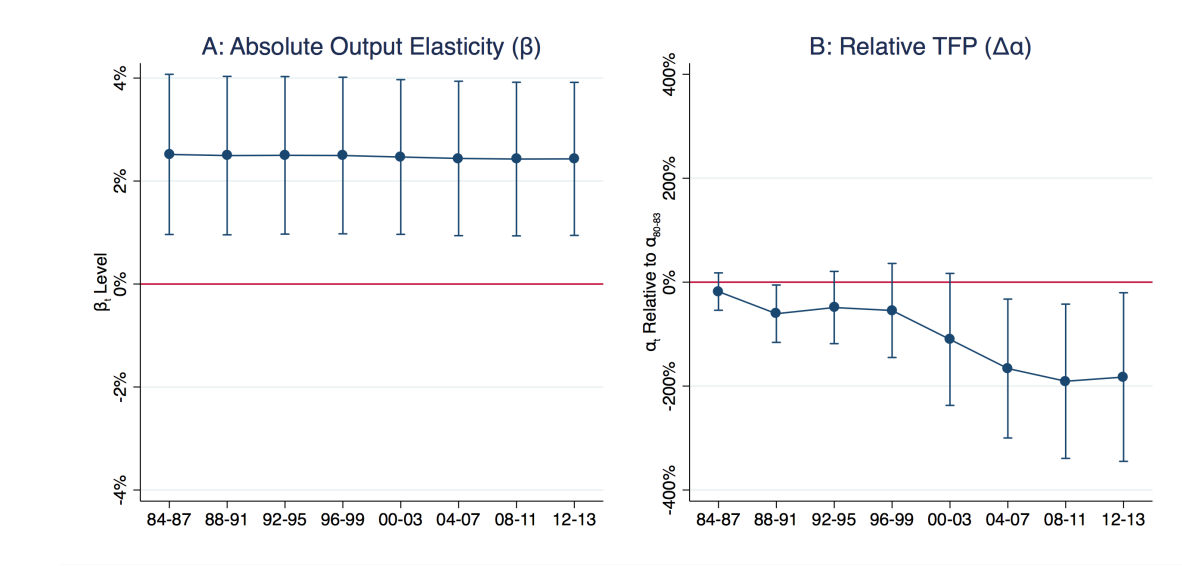
Note: Actual investments are based PhRMA annual reports. Predicted investments are based on the relationship between potential market size and R&D investments estimated with pre-2001 data assuming the market size-investment relationship is fixed.

Figure 2: Heterogeneity in the Elasticity of New Drugs with respect to Market Size



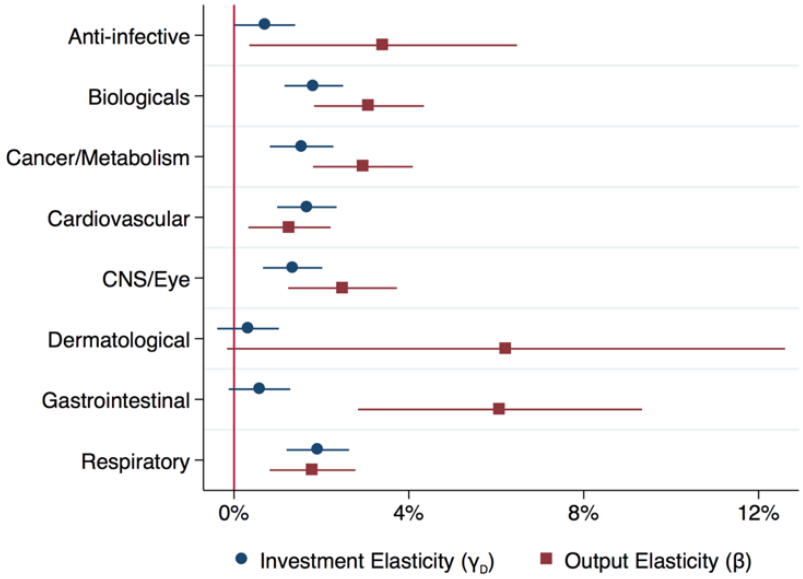
Note: Estimates are from two separate regressions based on Eq. 5.5, where the output elasticity parameter (β) is allowed to vary over time periods (Panel A) or R&D classes (Panel B).

Figure 3: Time-varying Estimates of Productivity Parameters



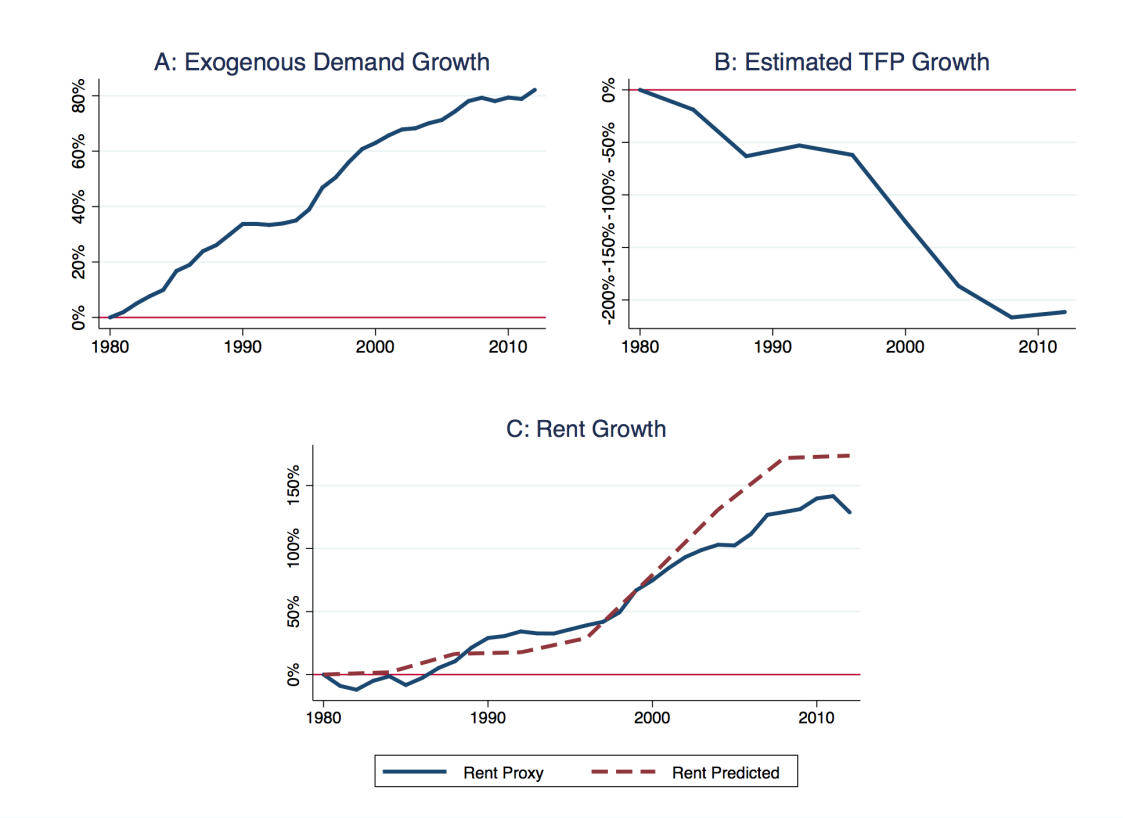
Note: Estimates are from two separate regressions based on the production function Eq. 5.5, where in Panel A the output elasticity parameter (β) is allowed to vary over time periods, and in Panel B the TFP parameter (α) is allowed to vary over time periods. Due to the conditional Poisson specification only relative changes in TFP can be estimated.

Figure 4: Heterogeneity of Investment and Production Elasticities



Note: Estimates are reported from two separate regressions, one based on the investment equation Eq. 3 and one on the production functions Eq. 5.5, in both cases where the focal parameter is allowed to vary over R&D classes.

Figure 5: Relating Demand Growth to Productivity and Rents



Note: All graphs display growth in measures relative to baseline estimates or data from 1980. Exogenous demand is per the demographics-based market size measure. TFP estimates are coefficient estimates re-reported from Figure 3, Panel B. Rent Proxy is constructed by revenues per market size per previously approved "on-patent" NME and Rent Predicted is the product of the demand growth and negative TFP growth. See Section 6 for the formulas for the Predicted and Proxy Rents, $\%\Delta\text{Rent}_y^1$ and $\%\Delta\text{Rent}_y^2$, respectively.

8 Appendix

8.1 Therapeutic class categorization

Table A1: Drug Category Crosswalk

Private R&D Category	**Matched ATC Category**
Anti-Infective	1. Anti-infective, non-viral [J01-J04] 2. Anti-viral [J05] 3. Parasitic [P]
Biological / Immunological	1. Biological / Immunological [L02-L04]
Cancer / Endocrine / Metabolism	1. Cancer [L01] 2. Diabetes / Obesity [A08, A10] 3. Hormonal [H] 4. Endocrine, reproductive [G03]
Cardiovascular	1. Blood [B] 2. Cardiovascular, non-blood-specific [C]
Central Nervous System / Eye	1. CNS [N] 2. Eye [S01]
Gastrointestinal / Genitourinary	1. Gastrointestinal [A01-A07, A09, A11-A15] 2. Kidney / Gynecological / Urological [G01, G02, G04]
Respiratory	1. Respiratory [R]
Dermatological	1. Dermatological / Musculoskeletal [D, M]

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.

8.2 Determination of private investment lags

Table A2: Private R&D lag tests

Lag (Years)	0	1	2	3	4	5	6
$\mathbb{E}[\text{Demand}]_{jy}$	2.260***	2.383***	2.471***	2.463***	2.476***	2.441***	2.533***
F-stat	43.6	47.7	57.6	61.1	75.0	92.4	136.6

Table A3: Private R&D lag tests, cont.'d

Lag (Years)	7	8	9	10	11	12
$\mathbb{E}[\text{Demand}]_{jy}$	2.588***	2.574***	2.688***	2.655***	2.647***	2.510***
F-stat	147.4	98.0	80.7	57.5	41.5	28.8

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