Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry*

Daron Acemoglu
MIT

Joshua Linn
MIT

June 2003.

Abstract

Is innovation primarily determined by scientific constraints and motivated by past scientific discoveries? Or does innovation respond to profit incentives? In this paper, we develop a simple model linking innovation to potential market size, and show that under a variety of circumstances, a greater market size for a particular product, which implies greater profitability from sales, spurs faster innovation for this product. We provide evidence for this hypothesis from the pharmaceutical industry by looking at changes in the potential market size for various drug categories driven by U.S. demographic changes. We find that a 1 percent increase in the potential market size for a drug category leads to a 4 to 7.5 percent increase in the number of new drugs in that category. This result is generally robust to using different estimation strategies, adopting different classifications of drugs into various categories, and controlling for pre-existing trends.

Keywords: Economic Growth, Entry, Innovation, Pharmaceuticals.
JEL Classification: O31, O33, L65.

*We thank David Autor, Ernie Berndt, Olivier Blanchard, Alfonso Gambardella, Amy Finkelstein, Stan Finkelstein, Jerry Hausman, Paul Joskow, and Frank Lichtenberg for useful comments. We are especially grateful to Frank Lichtenberg for generously sharing the Food and Drug Administration drug approval and the Computer Retrieval of Information on Scientific Projects data with us. This research is partially funded by the National Science Foundation and the Russell Sage foundation.

1 Introduction

This paper develops the theoretical and empirical case that market size is a key determinant of innovation. We construct a simple model linking innovation rates to current and future market size under a variety of circumstances, and provide evidence from the pharmaceutical industry to support this hypothesis. Our empirical work, which exploits changes in the market size for various drug categories driven by U.S. demographic trends, finds economically significant and relatively robust effects of market size on entry of new drugs.

Although many historical accounts of technological change focus on the autonomous progress of science and on major breakthroughs that take place as scientists build on each other's work,[1] economists typically emphasize incentives and the demand for innovation. John Stuart Mill provides an early statement of this view in his *Principles of Political Economy*, when he writes:

> "The labor of Watt in contriving the steam-engine was as essential a part of production as that of the mechanics who build or the engineers who work the instrument; and was undergone, no less than theirs, in the prospect of a renumeration from the produces" (1890, p. 68, also quoted in Schmookler, 1966, p. 210).[2]

The view that profit opportunities are the primary determinant of innovation and invention is most forcefully articulated by Griliches and Schmookler (1963), and especially by Schmookler in his seminal study, *Invention and Economic Growth*. Schmokler writes: "...invention is largely an economic activity which, like other economic activities, is pursued for gain" (1966, p. 206), and argues against the importance of major breakthroughs in science for economic innovation. He concludes from his analysis of innovations in petroleum refining, papermaking, railroading, and farming that there is no evidence that past breakthroughs have been the major factor in new innovations. He continues: "Instead, in hundreds of cases the stimulus was the recognition of a costly problem to be solved or a potentially profitable opportunity to be seized..." (1966, p. 199).

A main determinant of profitability of new innovations is the market size for the resulting product or technology. A greater market size increases profits and makes innovation and in-

[1] See, for example, Ceruzzi (2000), Rosenberg (1974) and Scherer (1984). Ceruzzi emphasizes the importance of a number of notable scientific discoveries and the role played by certain talented individuals in the development of modern computing. He points out how important developments took place despite the belief of many important figures in the development of the computer, such as Howard Aiken, that there would not be a market greater than a handful of personal computers in the United States (2000, p. 13).

[2] In fact, profits were very much in the minds of James Watt and his business partner Matthew Boulton in designing and marketing the steam-engine. Boulton wrote to Watt: "It is not worth my while to manufacture your engine for three countries only, but I find it very well worth my while to make it for all the world." (quoted in Scherer, 1984, p. 13). Similarly motivated by his profit motives, James Watt praised the patent system, arguing that: "...an engineer's life without patent was not worthwhile" (quoted in Mokyr, 1990, p. 248).

vention more desirable. To emphasize this point, Schmookler called two of his chapters "*The amount of invention is governed by the extent of the market.*" Schmokler's argument is most clearly illustrated by the example of the horseshoe. He documented that there was a very high rate of innovation throughout the late nineteenth and early twentieth centuries in the very ancient technology of horseshoe making, and no tendency for inventors to run out of additional improvements. On the contrary, inventions and patents increased because demand for horseshoes was high. Innovations came to an end only when "the steam traction engine and, later, internal combustion engine began to displace the horse..." (1966, p. 93).

The importance of profit incentives and market size in innovation is also essential for the recent endogenous technological change models, which make profit incentives the central driving force of the pace of aggregate technological progress (e.g., Aghion and Howitt, 1992, Grossman and Helpman, 1991, Romer, 1990), as well as to the induced innovation and directed technical change literatures, which investigates the influence of profit incentives on the types and biases of new technologies (see, for example, Kennedy, 1964, Drandkis and Phelps, 1965, Samuelson, 1965, Hayami and Ruttan, 1970, and Acemoglu, 1998, 2002, and 2003). A recent series of papers by Kremer, for example (2002), also build on the notion that pharmaceutical research is driven by market size to argue that there is generally insufficient research to develop cures for third-world diseases such as malaria.

There is relatively little convincing quantitative evidence, however, on the effect of profit opportunities on innovation, especially on the specific types of innovation, which is the central focus of the induced innovation and directed technical change literatures. The difficulty lies in part in finding a source of variation in profit opportunities associated with different types of technologies.

We believe that the pharmaceutical industry is a particularly appropriate place to look for the effects of profit motives and market size on innovation for several reasons. The pharmaceutical industry is one of the most innovative sectors of the economy, providing us with both an interesting and potentially representative sample of innovations. In addition, there are major changes in the market sizes for different types of drugs because of changes in the age composition of the U.S. population.[3]

Our empirical strategy is to exploit these variations in market size driven by demographic changes (or past demographic trends), which should be exogenous to other, for example scientific, determinants of innovation in drugs. We create the potential market size for various drug

[3]For a number of these drugs, non-U.S. markets may also be relevant. Nevertheless, the U.S. market is disproportionately important, constituting over 40 percent of the world market (IMS, 1999). Moreover, changes in the demographics of other major OECD economies are correlated with those in the United States, so our U.S.-based measure of potential market size will also capture these changes.

categories according to the age distribution of their users at a given point in time, and then trace changes in this potential market size driven by changes in demographics (holding the age profile of consumption of various drug categories constant over time). We measure innovation by the Food and Drug Administration's (FDA) approval of new drugs, which we match to our drug categories. These data are previously used by Lichtenberg and Virahbak (2002).

Our results show that there is an economically and statistically significant response of the introduction of new drugs to market size. For example, we find that a 1 percent increase in the size of the potential market for a drug leads to a 4 to 7.5 percent increase in the number of new drugs that are supplied to the market. New drugs that enter the market comprise both generics and non-generics. We find that both of these types of drugs respond to market size, but our results are somewhat weaker for non-generics. We further check the robustness of our results by controlling for trends in health insurance and the drug coverage of various insurance policies, lagged FDA approvals, pre-existing trends in innovation, differences in the non-economic incentives to innovate, and advances in biotechnology.

We also investigate whether it is the current market size or past or future market sizes that have the largest effect on entry of new drugs. On the one hand, because changes in demographics are known in advance, drug entry may respond to anticipated future market sizes. On the other hand, because the development process of new drugs can be long, entry may respond to past market sizes. In practice, we find that typically current market size and 5-10 years leads of market size have the strongest effect on entry rates of new drugs. We interpret this as evidence for limited anticipation effects, which are consistent with our theory.

There are a number of other studies related to our work. First, Schmookler (1966) documented a statistical association between investments and sales, on the one hand, and patents and innovation, on the other, and argued that the causality ran largely from the former to the latter. The classic study by Griliches (1957) on the spread of hybrid seed corn in U.S. agriculture also provides evidence consistent with the view that technological change and technology adoption are closely linked to profitability and market size. In more recent research, Pakes and Schankerman (1984) investigate this issue using a more structured approach linking R&D intensity at the industry level to factor demands and to growth of output. Their inter-industry results are highly supportive of Schmookler's conclusions. Scott Morton (1999) studies the decision of firms to introduce a new generic drug, and finds a positive correlation between entry into a new market and experience in a similar market as well as size of revenues in the target market. None of these studies exploit a potentially exogenous source of variation in market size, however, and do not establish a causal link from market size to innovation.

Second, some recent research has investigated the response of innovation to changes in energy prices. Most notably, Newell, Jaffee and Stavins (1999) show that between 1960 and 1980, before there was a reaction to the increase in energy prices, the typical air-conditioner sold at Sears became significantly cheaper, but not much more energy-efficient. On the other hand, between 1980 and 1990, there was little change in costs, but air-conditioners became much more energy-efficient. They argue that the technological developments in air-conditioning between 1980 and 1990 were a response to profit incentives created by the higher energy prices. In a related study, Popp (2002) also finds a strong correlation between energy prices, as well as existing scientific knowledge base, and aggregate patents. These findings are consistent with the hypothesis that the type of innovation responds to profit incentives, though they do not establish causality. Moreover, this evidence is relevant for the "price effect"—i.e., how innovation responds to factor price changes (see Acemoglu, 1998, and 2002)—not for the "market size effect", which is the focus here.

Third, there is substantial research focusing on innovations in the pharmaceutical industry, which are clearly related to the current paper. Henderson and Cockburn (1996), Cockburn and Henderson (2001), and Danzon, Nichelson and Sousa Pereira (2003) study the determinants of success in clinical trials, focusing mainly on firm and project size. Galambos and Sturchio (1998), Cockburn, Henderson and Stern (1999), Gambardella (2000), and Malerba and Orsenigo (2000) discuss various aspects of the recent technological developments in the pharmaceutical industry, including how firms have responded to new technological opportunities. Ling, Berndt and Frank (2003) investigate the complementarity between new technologies and the skills of physicians in the development and spread of new drugs. Lichtenberg (2003) presents evidence suggesting that the types of drugs that were developed changed towards drugs most useful for the elderly after Medicare was established. Finally, Lichtenberg and Waldfogel (2003) show that following the Orphan Drug Act there were larger declines in mortality among individuals with rare diseases (compared to other diseases), and interpret this as resulting from the incentives created by the Act to develop drugs for these rare diseases.

Most closely related to this study is Finkelstein (2002). She exploits three different policy changes affecting the reimbursement of costs of vaccination against 6 infectious diseases: the 1991 policy change that all infants be vaccinated against hepatitis B, the 1993 decision of Medicare to cover the costs of influenza vaccination, and the 1986 introduction of funds to insure vaccine manufactures against product liability lawsuits for vaccines against polio, diphteria, tetanus, measles, mumps, rubella, or pertussis. She finds that increases in vaccine profitability resulting from these policy changes are associated with a significant increase in the number of

clinical trials to develop new vaccines against the relevant diseases.

The rest of the paper is organized as follows. We outline a simple model linking innovation to market size in the next section. The analysis here also shows that the effect of market size on innovation generalizes to a variety of technological environments. Section 3 briefly explains our empirical strategy, and Section 4 describes the basic data sources and construction of the key variables. Sections 5 provides the central empirical results and a variety of robustness checks. Section 6 contain some concluding remarks, while the Appendix gives further data details.

2 Theory

In this section, we outline a simple framework for the analysis of the influence of market size on innovation. Subsection 2.1 builds a simple model where research can be directed to one of many potential drug types, and highlights the relationship between market size and innovation. In subsection 2.2, we discuss the implications of potential delays in the development and approval processes of new drugs. Subsection 2.3 generalizes the basic model to show the response of innovation effort and R&D to anticipated changes in future market size. Subsection 2.4 considers a generalization where research is only imperfectly directed, in the sense that research for a particular drug line may result in the discovery of better varieties of other drugs. Subsection 2.5 considers exogenous differences in technological drift across various drug lines, and finally subsection 2.6 extends the model to introduce an explicit distinction between entry of generic drugs and entry of non-generic drugs, which may more closely correspond to the common notion of "innovation".

2.1 Basic Model

Consider an economy consisting of a set I individuals, each denoted by i. Time is continuous $t \in [0, \infty)$, and all individuals are infinitely lived. There are two types of goods in this economy. First, a basic good, y, which can be consumed or used for the production of other goods, or for research expenditure. Individual i has an endowment $y_i(t)$ at time t. In this paper, we take these endowment processes as exogenous. Second, there is a large number J of drugs, x_1, ..., x_J. Each drug has a potentially time-varying "quality", $q_1(t)$, ..., $q_J(t)$. These qualities are improved by research and development as will be described below. Each individual demands only one type of drug. Hence, we partition the set I of individuals into J disjoint groups, G_1,...,G_J with $G_1 \cup G_2 \cup ... \cup G_J = I$, such that if $i \in G_j$, then individual i demands drug j.

More specifically, if $i \in G_j$, then his preferences are given by

$$\int_0^\infty \exp(-rt) \left[c_i(t)^{1-\gamma} \left(q_j(t) x_{ji}(t) \right)^\gamma \right] dt,$$

where r is the discount rate of the consumers (also the interest rate in the economy), $\gamma \in (0,1)$, $c_i(t)$ is the consumption of individual i of the basic good at time t, and $x_{ji}(t)$ is the consumption of individual i of drug j. This Cobb-Douglas functional form, which implies an elasticity of substitution equal to 1 between the basic good and drugs, and the assumption that each individual only consumes one type of drug are for simplicity and do not affect the results of interest here.[4]

Normalizing the price of the basic goods to 1 in all periods, and denoting the price of drug j at time t by $p_j(t)$, individual demands for drugs are given by

$$x_{ij}(t) = \begin{cases} \frac{\gamma y_i(t)}{p_j(t)} & \text{for } i \in G_j \\ 0 & \text{for } i \notin G_j, \end{cases} \tag{1}$$

for all $i \in I$ and for all $j = 1, ..., J$.

At any point in time, there is one firm with the best-practice technology for producing each type of drug. The best-practice firm in drug line j can produce one unit of drug with quality $q_j(t)$ using one unit of the basic good. Technological progress in this economy takes the form of increases in $q_j(t)$'s. If there is an innovation for drug line j currently with quality $q_j(t)$, this leads to the discovery of a new drug of quality $\lambda q_j(t)$ where $\lambda > 1$. For the purposes of the model, we think that any new innovation is approved (for example by the FDA) and can be sold to consumers immediately.

We start with a very simple formulation of the R&D technology whereby one unit of the final good devoted to R&D for drug line j leads to a flow rate of $\delta_j > 0$ of discovering a new drug of this type—or equivalently, if total R&D effort at time t is $z_j(t)$, the flow rate of innovation (the rate of entry of new drugs) for this line of drugs is $n_j(t) = \delta_j z_j(t)$. That the flow rates of innovation differ across drugs captures the possibility that technological progress is scientifically more difficult in some lines than others, which is the effect emphasized by science-driven theories of innovation discussed in the Introduction.

Notice the most important feature of this R&D technology for our focus here: technological progress is *directed* in the sense that firms can devote their research effort and expenditure to

[4] One implication of the Cobb-Douglas functional form is that the share of income that an individual spends on medicine is constant. This implication can be easily relaxed by considering a utility function with an elasticity of substitution different from 1, as in the factor market models with directed technical change (see, for example, Acemoglu, 1998, 2002, 2003b).

It is also straightforward to extend the model so that each individual demands potentially more than one type of drug, though this would require additional notation.

developing particular types of drugs. This contrasts with a different model where firms invest in R&D in an undirected way, and discover new versions of any one of a set of drugs. The pharmaceutical industry, especially in recent past, is the prime example of an industry where companies with fairly sophisticated R&D divisions or specialized R&D firms can undertake research for specific drug lines (e.g., Gambardella, 2000, Malerba and Orsenigo, 2000).[5]

The demand curves in (1) have an elasticity equal to 1, so the unconstrained monopolist would like to charge an arbitrarily high price. However, the firm with the best drug in line j is effectively competing with the next best drug in that line. Consider such a firm with quality $q_j(t)$ charging price $p_j(t)$. If this price is arbitrarily high, the next best quality could supply to the market and make positive profits, driving the best technology to zero profits. Therefore, the firm with the best drug has to set a *limit* price so as to exclude the next best firm, or, formally, make sure that consumers are happy to buy from it rather than buy from the next best firm even if the next best firm charges the lowest possible price, i.e., equal to its marginal cost, 1. Imagine the problem of a consumer in this case. If she buys from the best firm with quality $q_j(t)$ and price $p_j(t)$ and chooses her optimal consumption as given by (1), she will have instantaneous utility at time t equal to

$$\left(q_j(t)\right)^{\gamma}(1-\gamma)^{1-\gamma}\gamma^{\gamma}\left(p_j(t)\right)^{-\gamma}y_i(t),$$

and if she purchases from the next best firm, which, by definition, has quality $q_j(t)/\lambda$ and charges price equal to marginal cost, 1, she will have utility

$$\lambda^{-\gamma}\left(q_j(t)\right)^{\gamma}(1-\gamma)^{1-\gamma}\gamma^{\gamma}y_i(t).$$

The limit price sets these two expressions equal to each other, hence, equilibrium prices for all j and t satisfy:

$$p_j(t)=\lambda. \tag{2}$$

This implies that the instantaneous profits of the firm with best product (technology) with quality $q_j(t)$ in line j at time t is:

$$\begin{aligned}\pi_j\left(q_j(t)\right) &= (\lambda-1)\gamma\sum_{i\in G_j}y_i(t) \\ &= (\lambda-1)\gamma Y_j(t)\end{aligned} \tag{3}$$

[5]Naturally, there exist examples of research directed at a specific drug type leading to the discovery of a different product, such as the well-known example of Viagra, which resulted from research on hypertension and angina, and was partly accidentally discovered from the detection of side effects in a clinical study (see, e.g., Kling, 1998). Nevertheless, such examples appear to be the exception rather than the rule. We return to this issue below when we consider a hybrid model with both directed and undirected research, and also in the empirical part.

where the second line defines $Y_j(t) \equiv \sum_{i \in G_j} y_i(t)$ as the total income of the group of consumers demanding drug j, and corresponds to the market size for drug j. Throughout we assume that all $Y_j(t)$'s are known in advance, which is plausible in the context of demographically-driven changes in demand. This market size can change because the number of consumers demanding this product changes, or because their incomes change, or also possibly because new varieties of drugs steal consumers from this particular drug. Notice that profits of drug companies are independent from quality, $q_j(t)$, which is a feature of the Cobb-Douglas utility function.

All firms in this economy are owned by the consumers, are forward-looking, and discount future profits at the discount rate r. The discounted value of profits for firms can be written by a standard dynamic programming recursion. Denote the value of a firm that owns the most advanced drug of quality q_j in line j at time t by $V_j(t \mid q_j)$. This is given by:[6]

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \delta_j z_j(t \mid q_j) V_j(t \mid q_j) \tag{4}$$

for each $j = 1, 2, ..., J$, where $\pi_j(q_j)$ is the flow profits in drug line j given by (3), and $z_j(t \mid q_j)$ is equilibrium R&D effort at time t on this line by other firms when current technology is q_j (because of the standard replacement effect first emphasized by Arrow, 1963, the firm with the best technology does not undertake any R&D itself, see, for example, Aghion and Howitt, 1992). To simplify notation, we will typically use $z_j(t)$ instead of $z_j(t \mid q_j)$. Intuitively, the flow value of owning the best technology in line j, $rV_j(t \mid q_j)$, is equal to the flow profits, $\pi_j(q_j)$ plus the potential appreciation of the value, $\dot{V}_j(t \mid q_j)$, but also takes into account that at the flow rate $n_j(t) = \delta_j z_j(t)$ there will be a new innovation, thus the current firm will lose its best-practice status, and make zero profits thereafter.

For the purposes of the mapping this model into reality, note that entry of new drugs that are not technologically better, but steal customers from the incumbent, such as the entry of generics, can also be included in $z_j(t)$, especially since equation (3) shows that the profits that a new entrant makes is independent of the quality of its product, as long as it is sufficient to take over (part of) the market. Thus for now, we think of $z_j(t)$ and $n_j(t)$ as corresponding to the entry of both generic and non-generic drugs, and in subsection 2.6, we will present a model with separate entry rates of generics and non-generics.

There is free-entry into research and development to develop better quality drugs. Therefore, if there is positive research for some drug line $j = 1, 2, ..., J$ at time t, then the free-entry condition ensuring zero profits must hold. In other words,

$$\text{if } z_j(t) > 0, \text{ then } \delta_j V_j(t \mid q_j) = 1. \tag{5}$$

[6] Throughout, we assume that the relevant transversality conditions hold and discounted values are finite.

Alternatively, we might have $z_j(t) = 0$, and $\delta_j V_j(t \mid q_j) \leq 1$, in which case research is not profitable, and in equilibrium, there will be no innovation.

An equilibrium in this economy is sequences of prices $p_j(t)|_{j=1,..J}$ that satisfy (2), consumer demands for drugs $x_i(t)|_{i \in I}$ that satisfy (1) and a sequence of R&D levels $z_j(t)|_{j=1,..J}$ that satisfy (5) with $V_j(\cdot)$ given by (4).

An equilibrium is straightforward to characterize. Differentiating equation (5) with respect to time implies that we must always have $\dot{V}_j(t \mid q_j) = 0$ for each $j = 1, 2, ..., J$ as long as $z_j(t) > 0$. Substituting this equation and (5) into (4) yields:

$$z_j(t) = \max\left\{\frac{\delta_j(\lambda - 1)\gamma Y_j(t) - r}{\delta_j}; 0\right\}, \tag{6}$$

for each $j = 1, 2, ..., J$, and for all t. From now on, unless otherwise stated, we assume that $Y_j(t)$'s are such that all equilibrium research levels are strictly positive, i.e., $z_j(t) > 0$ for all j and t, so that $z_j(t) = \frac{\delta_j(\lambda-1)\gamma Y_j(t)-r}{\delta_j}$, and we will often drop the max operator.

We now summarize the equilibrium allocation for future reference:

Proposition 1: In the economy described above, there exists a unique equilibrium in which R&D level in each drug line $j = 1, 2, ..., J$, $z_j(t)$, is given by (6) for all t.

The most important feature of (6) is that it highlights the market size effect in innovation, which is the main focus of this paper. The greater is the market size for a particular drug, the more profitable it is to be the supplier of that drug, and thus there will be greater research effort to acquire this position. Our empirical work below will investigate the strength of this effect in the pharmaceutical industry over recent decades. In addition, naturally, a higher productivity of R&D as captured by δ_j also increases R&D, and a higher interest rate reduces R&D since current R&D expenditures are rewarded by future revenues.

Another important implication of this equation is that there are no transitional dynamics. At any point in time, the amount of effort devoted to developing a particular drug line is determined by the current market size. Past market sizes and anticipated future market sizes do not affect current research effort. This is an implication of the linear R&D technology, which ensures that whenever there are profit opportunities, there will immediately be enough R&D to arbitrage them, thus ensuring $\dot{V}_j(t \mid q_j) = 0$. The intuition for the lack of a response to anticipated changes in future market size here highlights an important effect in quality ladder models of technological progress: with a greater market size in the future, firms would like to own the best-practice product at the time when the market size has actually become larger. Investing in R&D in advance could be useful to the extent that it achieves this objective.

However, it is not beneficial to invest in R&D too much in advance, since some other firm would improve over this innovation by the time the new and larger market size materializes. With the linear model here, z_j can change discontinuously, so investing even a little bit in advance of the actual increase in the size of the market is not profitable.[7] In subsection 2.3, we will look at a more general technology, which will introduce responses to anticipated changes in market size, but the same reasoning will limit the extent of these responses.

Also note that equation (6) provides us with a relationship we can easily take to the data. In fact, in this model research, $z_j(t)$, will have a success rate of $n_j(t) = \delta_j z_j(t)$, thus we can write:

$$n_j(t) = \delta_j(\lambda - 1)\gamma Y_j(t) - r. \tag{7}$$

This equation relates research output (innovation or entry), which we will approximate with FDA approval of new drugs, to market size, in this case the total expenditure of consumers in this line of drug. Since such expenditures are potentially endogenous, for example, depending on drug prices, we will try to exploit sources of exogenous variations originating from demographic changes. In addition, this equation also encapsulates the alternative view of the determinants of innovation, which maintains that cross-drug distribution of R&D is determined largely by technological research opportunities or perhaps by other non-profit related motives. If there are large and potentially time-varying differences in δ_j's, then these may be the primary factor determining variations in R&D across drug lines, and the market sizes may not have an important effect. Whether this is so or not is an empirical question.

2.2 Delays in Development and Approval

The baseline model ignores the potential delays in the process of development and approval of new drugs (for example, DiMasi et al., 1991, report that the eventual marketing of a drug may take as much as 15 years from the beginning of initial research). To incorporate such delays in the simplest possible way, suppose that it takes an interval of length T after the research decision for the drug to be developed, gain approval, and enter the market.

Given this structure, the key value function becomes:

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \delta_j z_j(t-T) V_j(t \mid q_j) \tag{8}$$

instead of (4) above, where $z_j(t-T)$ is the rate of innovation at time $t-T$. Equation (8) is a delayed differential equation rather than an ordinary differential equation, so a general analysis

[7] In practice, companies may also have an incentive not to market their discoveries before the market size increases in order to prevent competitors from leapfrogging their new product.

is more difficult. Nevertheless, the unique equilibrium in this case is still easy to characterize because of the simple structure here. To do this, note that the free entry condition now changes to

$$\text{if } z_j(t-T) > 0, \text{ then } \exp(-rT)\,\delta_j V_j(t \mid q_j) = 1, \tag{9}$$

which recognizes that innovation effort at time $t-T$ will lead to revenues at time t, hence the discounting for the interval of length T. Equations (8) and (9) together imply that:

$$z_j(t-T) = \max\left\{\frac{\exp(-rT)\,\delta_j(\lambda-1)\,\gamma Y_j(t) - r}{\delta_j}; 0\right\}, \tag{10}$$

which is very similar to (6), except for the term $\exp(-rT)$. This term takes into account that because of the development and approval delays, costs of R&D are incurred before the benefits accrue. This equation also makes it clear that longer development and approval delays discourage innovation.

Equation (10) may give the impression that there should now be a stronger response of innovation to future market sizes. This is not the case, however, since what we measure in the data is not the actual R&D expenditure, but entry of new drugs. In this model, entry of new drugs in category j at time t will be given by $n_j(t) = \delta_j z_j(t-T)$, thus the key prediction of the model changes to (ignoring the max operator):

$$n_j(t) = \exp(-rT)\,\delta_j(\lambda-1)\,\gamma Y_j(t) - r,$$

which only differs from (7) because of the term $\exp(-rT)$. This analysis therefore shows that delays in development and approval processes do not change the basic predictions of the theory.

2.3 Anticipation Effects

The baseline model considered the equilibrium with a simple linear R&D technology, which implied no transitional dynamics, and more importantly, no response to anticipated changes in future market sizes. We now generalize this basic setup in a simple way to obtain a reaction to (anticipated) future market sizes. To do this, we change the baseline model in one dimension: we assume that one unit of final good spent for R&D in line j leads to the discovery of a better drug at the flow rate $\delta_j z_j \phi(z_j)$, where z_j is the aggregate research effort devoted to the discovery of a new drug in this line. We also assume that $\phi'(z) \leq 0$ for all z, which implies that greater research effort runs into decreasing returns within a given period (there are constant returns to scale when $\phi(z) = 1$ for all z), but throughout $z\phi(z)$ is strictly increasing in z, so that greater aggregate research effort always leads to faster innovation in the aggregate.

Finally, free-entry into R&D for all lines implies that any new firm can enter taking $z_j(t)$ as given, thus without taking into account the reduction that its entry causes in the innovation rates of other firms.[8]

Given this specification, the value function changes from (4) to:

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \delta_j z_j(t) \phi(z_j(t)) V_j(t \mid q_j), \tag{11}$$

for each $j = 1, 2, ..., J$, which only differs from (4) because the flow rate of innovation is now $\delta_j z_j(t) \phi(z_j(t))$ rather than $\delta_j z_j(t)$.

Since each potential entrant takes the aggregate research effort in each line of drug as given, it anticipates that one unit of the basic good spent for R&D in drug line j will lead to an innovation at the flow rate $\delta_j \phi(z_j(t))$. Thus, the free-entry condition is

$$\delta_j \phi(z_j(t)) V_j(t \mid q_j) = 1, \tag{12}$$

for each $j = 1, 2, ..., J$ (again as long as $z_j(t) > 0$).

An equilibrium is defined similar to before, except that now the sequence of R&D levels $z_j(t)|_{j=1,...J}$ have to satisfy (12) instead of (5) with $V_j(\cdot)$ given by (11).

To make further progress in this case, let us assume that $Y_j(t) = Y_j$ for all t, in other words, that market sizes for different drugs are not changing over time. Then differentiate (12) with respect to time, which yields

$$\dot{V}_j(t \mid q_j) = \frac{\varepsilon_\phi(z_j(t))}{\delta_j \phi(z_j(t))} \frac{\dot{z}_j(t)}{z_j(t)},$$

where $\varepsilon_\phi(z_j(t)) = -\phi'(z_j(t)) z_j(t) / \phi(z_j(t))$ is the elasticity of the ϕ function. Substituting this and (12) into (11) and dividing by $\delta_j \phi(z_j(t))$ for each j, we obtain J differential equations in $z_j(t)$'s:

$$\frac{\dot{z}_j(t)}{z_j(t)} = \frac{1}{\varepsilon_\phi(z_j(t))} \left[r + \delta_j z_j(t) \phi(z_j(t)) - \delta_j \phi(z_j(t)) (\lambda - 1) \gamma Y_j \right]. \tag{13}$$

Since consumer incomes are now assumed to be constant, the steady-state equilibrium must have $\dot{z}_j(t) = 0$, which gives

$$z_j^S = \frac{\delta_j \phi\left(z_j^S\right) (\lambda - 1) \gamma Y_j - r}{\delta_j \phi\left(z_j^S\right)} \tag{14}$$

[8] This is the natural assumption given free entry. The alternative would be to assume that there is a consortium of firms in each line, jointly maximizing profits. In this case, the free entry condition below would change to: $\delta_j \left[\phi(z_j(q_j(t))) + z_j(q_j(t)) \phi'(z_j(q_j(t))) \right] V_j(q_j(t)) = 1$. This does not affect any of the results of the analysis.

as the steady-state R&D level in drug line $j = 1, 2, ..., J$. In the special case where $\phi(.) = 1$, this equation boils down to (6) from the previous subsection. Moreover, the steady-state R&D levels have the same features as the equilibrium there: a greater market size increases R&D, a greater δ_j, which corresponds to better research opportunities for this drug line, increases R&D, and higher interest rates reduce R&D.

An important feature of this equation, which follows from the simplifying assumptions we employed, is that the equilibrium behavior of $z_j(t)$ is independent of research and profitability in other drug lines. This feature simplifies the dynamics substantially. In addition, the right hand side of (13) is strictly increasing in $z_j(t)$ whenever $z_j(t) = z_j^S$, which implies that there can be at most one intersection of the right hand side with the 0 axis, and at this point of intersection, $\dot{z}_j(t)/z_j(t)$ is increasing in $z_j(t)$, as drawn in Figure 1. Therefore, (13) defines an unstable differential equation. This implies that starting away from the steady-state, the equilibrium $z_j(t)$ has to immediately jump to its steady-state value as given by (14). Hence, there are no transitional dynamics in this extended model either.

However, there is now an equilibrium response to anticipated future changes in market size. Consider the following situation: it is suddenly announced at date t' that Y_j will increase to $\hat{Y}_j$ in some future date $\hat{t} > t'$. How will this fully-anticipated change in market size affect equilibrium R&D? Suppose there is no change in $z_j(t)$ until $\hat{t}$. This implies that $z_j(t)$ has to jump up discontinuously at $t = \hat{t}$. But this implies that anticipating this jump, $V_j(t \mid q_j)$ will be changing before $\hat{t}$, in particular, $\dot{V}_j(t \mid q_j) < 0$. Since $z_j(t)$ is constant, this would violate the free-entry condition, (12). This reasoning implies that there should be no anticipated jumps in $z_j(t)$, in particular no jump at $t = \hat{t}$. This is only possible if $z_j(t)$ jumps by a small amount initially at $t = t'$, and then smoothly increases towards the new steady-state equilibrium. Therefore, in this model there are no transitional dynamics starting away from the steady-state, but R&D responds to anticipated future market size changes. Nevertheless, the same considerations as in the previous subsection, that increasing R&D investment too far in advance would not be profitable because somebody else would innovate over the new products before the actual increase in market size materializes, limit this response. In other words, even if a large change in market size is anticipated far in advance, we would not expect a large response long time in advance, but the response to build up gradually. In terms of our empirical work, even if demographic changes are anticipated at least 20 or 30 years in advance, we may expect significant innovation responses much later, perhaps 5 or 10 years in advance or even contemporaneously.

The next proposition summarizes the results and compares directly with Proposition 1.

Proposition 2: Consider the economy described in this subsection with $Y_j(t) = Y_j$. Then there exists a unique steady-state equilibrium in which R&D level in industry j, z_j^S, is given by (14).

If $z_j(t) \neq z_j^S$, then the equilibrium R&D level $z_j(t)$ immediately jumps to z_j^S.

If there is a change at time $t = t'$ such that it becomes common knowledge that $Y_j(t) = Y_j$ for $t < \hat{t}$ and $Y_j(t) = \hat{Y}_j > Y_j$ for all $t \geq \hat{t}$ where $\hat{t} > t'$, then $z_j(t')$ immediately jumps above z_j^S, and smoothly converges to $\hat{z}_j^S$, reaching it exactly at $t = \hat{t}$, where $\hat{z}_j^S$ is the new steady-state value of R&D level in drug line j for $Y_j(t) = \hat{Y}_j$ for all t.

2.4 Robustness: Imperfectly Directed Research

The analysis so far assumed that research was perfectly directed in the sense that a firm could explicitly undertake research for drug line j, which would lead to innovations only in this particular drug line. Given the randomness involved in scientific research, a more reasonable alternative might be to presume that research is only imperfectly directed. To investigate the theoretical implications of this, we return to the baseline model of subsection 2.1, and assume that one unit of the basic good spent for R&D in the drug line j leads to a flow rate of $p\delta_j$ of discovering a new drug of this type, and to a flow rate of $\frac{1-p}{J}\delta_{j'}$ any $j' = 1, ..., J$, where $1 \geq p > 0$. This technology implies that the parameter p captures the degree to which research is directed. When $p = 1$, we have the model of subsection 2.1, while with $p \to 0$, research becomes undirected.

With this change in research technology, the dynamic programming recursion for the value function changes to

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \left[pz_j(t) + (1-p)\sum_{j'=1}^{J} z_{j'} \right] \delta_j V_j(t \mid q_j), \tag{15}$$

for each $j = 1, ..., J$. This value function takes into account that a fraction $1-p$ of the research directed at other drug lines will result in innovation for this particular drug, thus the rate of innovation for drug line j is

$$n_j(t) = \left[pz_j(t) + (1-p)\sum_{j'=1}^{J} z_{j'} \right] \delta_j.$$

The free-entry condition also changes for the same reason, and becomes:

$$p\delta_j V_j(t \mid q_j) + (1-p)\sum_{j'=1}^{J} \delta_{j'} V_{j'}(t \mid q_{j'}) = 1 \tag{16}$$

for each j.

An equilibrium is defined similarly to before, except that now the sequence of R&D levels $z_j(t)|_{j=1,...J}$ have to satisfy (16) with $V_j(\cdot)$ given by (15).

Summing (16) over j, we obtain that $\sum_{j=1}^{J} \delta_j V_j(t \mid q_j) = 1$, and using this, (16) implies

$$\delta_j V_j(t \mid q_j) = 1$$

for each $j = 1, ..., J$, which also immediately implies $\dot{V}_j(t \mid q_j) = 0$. Now substituting these equilibrium conditions back into (15), and solving the resulting J equations simultaneously, we obtain

$$z_j(t) = \frac{1}{p}\left[(\lambda - 1)\gamma Y_j(t) - \frac{r}{\delta_j} - (1-p)\left((\lambda-1)\gamma \sum_{j'=1}^{J} Y_{j'}(t) - \sum_{j'=1}^{J} \frac{r}{\delta_{j'}}\right)\right]. \tag{17}$$

The interesting implication of this equation is that even when p is very small (i.e., $p \to 0$), there is a very strong market size effect linking research directed to drug line j to the market size and to profits from sales of that drug. In fact, equation (17) shows the opposite of an ad hoc intuition: the response of $z_j(t)$ to the market size for drug line j becomes stronger, not weaker, when p is lower. To understand this potentially paradoxical result, notice that directing research to a drug line that has a larger market size becomes less profitable because some of this research will lead to discoveries in other fields, which do not have correspondingly larger market sizes. But on the other hand, the fact that much of the research directed to this drug line in the future will be dissipated also means that rents from producing the best drug for a particular line will persist for longer. This second effect dominates, thus $z_j(t)$ is now more sensitive to the market size of drug j.[9]

Nevertheless this result is somewhat misleading since in this model with imperfectly directed research, what matters more, and what we will observe in practice, is innovation rates in a particular drug line, the $n_j(t)$'s, and using (17), we have:

$$n_j(t) = \delta_j(\lambda - 1)\gamma Y_j(t) - r, \tag{18}$$

which is identical to (7), and shows that this framework leads to similar results even when research is imperfectly directed.

Proposition 3: In the economy described above for any $p \in (0, 1]$, there exists a unique equilibrium in which R&D level in each drug line $j = 1, 2, ..., J$, the equilibrium R&D

[9]The reason why the first effect is dominated by the second is that given the linear R&D technology, in equilibrium we have $\delta_j V_j(t \mid q_j) = 1$ at all points in time for all j, so the fact that some of the research directed at drug line j leads to discoveries of other drugs does not discourage research towards j.

level, $z_j(t)$, is given by (17) for all t, and the equilibrium innovation rate, $n_j(t)$, is given by (18) and is invariant to the level of p.

This proposition establishes that the basic result derived in this section is robust to generalizing our setup to an environment where research is only partially directed. This result reflects the very strong tendencies towards no profits (no arbitrage) in each product (or technology), present in the endogenous technological change models based on quality ladders. In the next subsection, we will see that the same tendencies ensure that the results are also robust to introducing random discoveries at different rates in each drug line.

2.5 Robustness: Technological Drift

So far we have assumed that all innovation results from profit incentives, even if not all research may be directed perfectly as in the previous subsection. Historical accounts of many discoveries point out that there could also be a large amount of "technological drift", meaning random innovations arising from non-profit and non-economic motives. Even though we have reasons to suspect that such technological drift is limited in the pharmaceutical industry, where research is highly organized and very resource-intensive because of the necessary clinical trials and the FDA approval process, it is useful to briefly investigate the implications of this type of technological drift. The basic result of this subsection is that the presence of technological drift does not affect the form of the equilibrium derived above.

To incorporate technological drift into our framework, assume that innovations in drug line j at time t are now given by

$$n_j(t) = \delta_j z_j(t) + \xi_j(t) \tag{19}$$

where $\xi_j(t) \geq 0$ is the drift term, while $z_j(t)$ is R&D level as before. Let us assume that the time sequence of $\xi_j(t)$ is perfectly known by all agents (which is, naturally, a strong assumption). Then, the dynamic programming recursion for the value of owning the best product in drug line $j = 1, 2, ..., J$, changes to

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(t \mid q_j) - \left(\delta_j z_j(t) + \xi_j(t)\right) V_j(t \mid q_j).$$

Similar analysis to before implies that as long as there is research in a particular drug line $j = 1, 2, ..., J$ at time t, we must have the same free-entry condition as before, i.e.,

$$\text{if } z_j(t) > 0, \text{ then } \delta_j V_j(t \mid q_j) = 1,$$

but also

$$\text{if } z_j(t) = 0, \text{ then } \delta_j V_j(t \mid q_j) \leq 1$$

is also consistent with equilibrium, and results when $\xi_j(t)$ is so high that research in this drug line at this time would be a loss-making endeavor.

Combining these equations, we find that the unique equilibrium has

$$z_j(t) = \max\left\{\frac{\delta_j(\lambda-1)\gamma Y_j(t) - r - \xi_j(t)}{\delta_j}; 0\right\}, \tag{20}$$

for each $j = 1, 2, ..., J$, and for all t, and therefore, the overall amount of innovation is given by

$$n_j(t) = \max\left\{\delta_j(\lambda-1)\gamma Y_j(t) - r; \xi_j(t)\right\}. \tag{21}$$

The implications of this extended model are therefore identical to our baseline model, unless technology drift is so important that the amount of non-economic innovation exceeds the equilibrium innovation that would have resulted in the absence of such technological drift. In that case, innovation rates would be determined only technological drift, or by purely non-economic factors as conjectured by the extreme science-driven views of innovation. Once again, whether this is the case or not is an empirical question, which we turn to next.

2.6 Generics and Non-Generics

The analysis so far did not distinguish between generics and non-generics. In the empirical work, we will first look at all new entries, but then distinguish between generics and non-generic drugs—which likely better correspond to "innovation". We now briefly discuss how the predictions of the model change when we incorporate a distinction between generics and non-generics. We do this in the simplest possible way, and assume that pharmaceutical firms can engage in R&D to discover new drugs as described above, in particular in subsection 2.1, or they can engage in costly development to introduce a generic version of an already-existing drug (without quality improvement). Although bringing a generic drug to the market does not involve original research, it still requires substantial resources spent upfront (for the approval process or for marketing).[10]

Let us suppose that one unit of the final good devoted to prepare a generic for the market in drug line j leads to the successful entry of the generic at time t at the flow rate θ_j, and denote total generic development expenditure for drug line j at time t by $x_j(t)$. In practice, patent life also influences θ_j, since a longer patent can be approximated by a lower θ_j. Throughout we assume that $\theta_j > \delta_j$, since introducing a generic to the market must be easier than inventing a new drug.

[10] An alternative approach would be to link the entry of non-generics directly to the delayed entry of generics, since generics become possible only after the patents on previous non-generics expire. Nevertheless, there are significant profit incentives here, since to introduce a generic into the market is not a costless endeavor. A hybrid model incorporating both such delays and profit incentives is significantly more complicated to analyze.

If there is entry of a generic into drug line j at time t, we assume that both the incumbent and the generic entrant receive profits of $\mu(\lambda-1)\gamma Y_j(t)$ where $Y_j(t)$ is defined in (3) above as the market size for drug j at time t, and $\mu \in [0, 1/2)$. Recall that if the generic entrant and the incumbent engage in Bertrand competition, then they will both charge marginal cost, and we would have $\mu = 0$. The formulation here allows some degree of non-Bertrand competition (e.g., Cournot) or collusion, so $\mu > 0$ is possible. If $\mu = 0$, there would be no entry of generics, and the results in subsection 2.1 would apply.

In addition, we assume that in a market that already includes the incumbent and a generic producer, there is no further room for a third producer, so we can ignore the potential entry from further generic producers. Allowing further entry of this sort introduces additional notation, but does not affect the qualitative results.

Given this structure, the value of innovation (for a non-generic) is now given by

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \delta_j z_j(q_j) V_j(t \mid q_j) - \theta_j x_j(t) [V_j(t \mid q_j) - W_j(t \mid q_j)], \quad (22)$$

where $\pi_j(q_j) = \mu(\lambda-1)\gamma Y_j(t)$ as before and $W_j(t \mid q_j)$ is the value of being one of two producers supplying drug j at time t. The main difference between this expression and (4) above is the last term, which takes into account that at the flow rate $\theta_j x_j(t)$ there will be entry of a generic, in which case the innovator loses its monopoly position, and the associated value $V_j(t \mid q_j)$, and becomes one of two producers, receiving value $W_j(t \mid q_j)$.

With a similar reasoning to before, the dynamic programming recursion for the value of being one of two producers in the market is given by:

$$rW_j(t \mid q_j) - \dot{W}_j(t \mid q_j) = \mu\pi_j(q_j) - \delta_j z_j(t) W_j(t \mid q_j). \quad (23)$$

Intuitively, the only reason why the flow of profits captured by $W_j(t \mid q_j)$ will come to an end is because there is a better drug introduced to the market, which happens at the rate $\delta_j z_j(t)$.

Free entry requires that

$$\begin{aligned} z_j(t) &\geq 0, \text{ and } \delta_j V_j(t \mid q_j) \leq 1 \text{ with complementary slackness} \\ x_j(t) &\geq 0, \text{ and } \theta_j W_j(t \mid q_j) \leq 1 \text{ with complementary slackness,} \end{aligned}$$

which is similar to (5) above, but we have written it explicitly in the form of a complementary slackness condition to emphasize that there may not be R&D for new drugs or any generic entry under certain conditions.

Let us the next assume that

$$\mu\theta_j < \delta_j \quad (24)$$

for all j. If this assumption does not hold, the entry of new generics is so profitable and sufficiently rapid that it ceases to become profitable for pharmaceutical companies to undertake R&D for introducing new drugs, and consequently, quality improvements come to an end. As long as Assumption (24) holds, similar arguments to before imply that the unique equilibrium is given by:

$$\begin{aligned} z_j(t) &= \frac{\mu\theta_j(\lambda-1)\gamma Y_j(t) - r}{\delta_j}, \\ x_j(t) &= \frac{(\delta_j - \mu\theta_j)(\lambda-1)\gamma Y_j(t)}{\theta_j - \delta_j}. \end{aligned} \tag{25}$$

In contrast, it can be verified that if $\mu = 0$, the equilibrium of Proposition 1 applies, and if Assumption (24) does not hold, there will be no R&D, i.e., $z_j(t) = 0$, and thus $\lim_{t\to\infty} x_j(t) = 0$, as they will eventually be two producers in the market for drug j.

Moreover, let $n_j(t) = \delta_j z_j(t)$ be the the entry rate of new drugs, and $g_j(t) = \theta_j x_j(t)$ be the entry rate of generics. Then, we also have:

$$\begin{aligned} n_j(t) &= \mu\theta_j(\lambda-1)\gamma Y_j(t) - r, \\ g_j(t) &= \frac{\theta_j(\delta_j - \mu\theta_j)(\lambda-1)\gamma Y_j(t)}{\theta_j - \delta_j}. \end{aligned} \tag{26}$$

There are a number of important points to note about this equilibrium. First, both $n_j(t)$ and $g_j(t)$, i.e., the entry rates of both non-generics and generics, respond positively to market size. Therefore, this model generalizes the key prediction of our baseline model in subsection 2.1. Second, other comparative static results are now quite different than in the baseline model. The entry rates of non-generics no longer respond to δ_j (and $z_j(t)$ is decreasing in δ_j). Instead, they respond positively to μ and θ_j (two parameters that should intuitively make entry of generics more profitable). This is because the rate of entry of non-generics is determined to ensure 0 profits for generics; once generic drugs enter the market, their producers will continue to make profits until there is a new and better drug. Finally, differentiating the equations in (26) with respect to $Y_j(t)$ shows that either generics or non-generics may respond more to changes in market size (it depends on whether μ or $(\delta_j - \mu\theta_j)/(\theta_j - \delta_j)$ is larger). Plausibly, we expect μ to be small and $\theta_j - \delta_j$ to be large, and therefore, generic entry to be more responsive to changes in market size than non-generic entry.[11]

[11] In practice, the presence of variable development and approval for non-generics may also imply a smaller response to the current market size for these drugs.

3 Empirical Strategy

3.1 Estimation Issues

We are interested in testing whether an increase in the market size, and hence the profitability, of a particular category of drugs is associated with an increase in innovation in that category, as suggested by the theoretical relationship in equation (7). As $r \to 0$, we can take logs on both sides of this equation to obtain:

$$\log n_j(t) = \text{constant} + \log \delta_j + \log m_j(t), \tag{27}$$

where $m_j(t) \equiv \gamma Y_j(t)$ is the market size for drug line j at time t. We will proxy innovation or entry of new drugs, $n_j(t)$, by new drug approvals by the FDA (Food and Drug Administration) in broad drug categories as described below. This measure, denoted by N_{ct} for drug category c at time t, could include the entry of generic drugs.[12] Although generic drugs do not correspond to "innovation" according to the standard usage of this term, they are still driven by the same profit incentives as innovation, and are similar to innovation in the context of our model—as noted above, they steal customers from the incumbent. After presenting results using all drug approvals, we separate generics from non-generics, and investigate whether the relationship between market size and entry differs for the two types of drugs. Instead of actual market size, $m_j(t)$, we will use potential market size driven by demographic changes, which we denote by M_{ct}, and discuss its construction below. The $\log \delta_j$ terms correspond to fixed effects in this equation.

Therefore, adding other potential determinants and an error term capturing other unobserved influences, and allowing the coefficient of $\log M_{ct}$ to differ from 1 as it will do with more general preferences than Cobb-Douglas, we arrive at an estimating equation of the form:

$$\log N_{ct} = \alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \varepsilon_{ct}, \tag{28}$$

where N_{ct} is the number of new drugs in category c in time period t, M_{ct} is potential market size, X'_{ct} is a vector of controls, including a constant, with β as the corresponding vector of coefficients, and ε_{ct} is a random disturbance term, capturing all omitted influences. The dependent variable is the logarithm of new drugs, so that other factors lead to proportional changes in the entry of new drugs. In addition, we will also experiment with models that have leads and lags of $\log M_{ct}$ in order to investigate delays and anticipation effects in the R&D process.

[12] Two other potential proxies for innovation rates would be patent rates and data on clinical trials. Patent data, which exist at the detailed industry level (see, e.g., Jaffe, Trachtenberg and Henderson, 1996), does not have enough information to match to detailed drug categories. We were also unable to obtain data on clinical trials for a sufficient number of drug types.

One problem is that N_{ct} is a count variable (number of new drugs), so it can equal 0. In our data, this is not common, but in most specifications there are typically a few drug category-time cells where N_{ct} is equal to 0. This makes the estimation of (28) impossible. We take a number of approaches to this problem.

First, we change (28) to

$$\log \tilde{N}_{ct} = \alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \gamma \cdot d_{ct} + \varepsilon_{ct}, \tag{29}$$

where $\tilde{N}_{ct} = N_{ct}$ if $N_{ct} \geq 1$ and $\tilde{N}_{ct} = 1$ if $N_{ct} = 0$, and the variable d_{ct} is a dummy that equals 1 when there are no approvals, i.e., $d_{ct} = 1$ if $N_{ct} = 0$ and $d_{ct} = 0$ otherwise. This procedure was used by Pakes and Griliches (1980) and has the advantage of simplicity and flexibility (the data determine how $N_{ct} = 0$ should be treated). The drawkback of this procedure is that the variable d_{ct} is mechanically a function of N_{ct}, so it can introduce various forms of biases.

Perhaps more satisfactory is to consider the following Poisson model (see, for example, Wooldridge, 1999, 2002, or Hausman, Hall, and Griliches, 1984):

$$N_{ct} = \exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta) + \varepsilon_{ct}, \tag{30}$$

which can be seen as a slight variant on equation (27) above. Our results below will show that the two models give similar results when there are only a few empty approval cells. When there are more empty cells, as when we look separately at generics and non-generics, there can be somewhat larger discrepancies between the two models, however. In those cases we favor the Poisson model.

There are likely to be significant differences in the technology of R&D and entry of new drugs across drug categories, and potentially over time. For this reason, throughout we will include a full set of drug category and time dummies.[13] Thus our baseline linear regression model is:

$$\log N_{ct} = \alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \gamma \cdot d_{ct} + \zeta_c + \mu_t + \varepsilon_{ct}, \tag{31}$$

where ζ_c denotes a full set of drug category dummies, and μ_t a full set of time dummies. The corresponding Poisson model would be:

$$N_{ct} = \exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \zeta_c + \mu_t) + \varepsilon_{ct}. \tag{32}$$

However, the estimation of (32) would lead to biased estimates, since the fixed effects ζ_c cannot be estimated consistently. To deal with this problem, we follow Hausman, Hall, and Griliches

[13] This is the reason why we do not want to estimate (7) directly and use either (28) or (29); time dummies should have proportional effects on entry of new drugs, whereas a level regression, as in (7), would force them to have level effects, biasing inference.

(1984), and transform (32) to obtain:

$$S_{ct} = \frac{\exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \mu_t)}{\sum_{\tau=1}^{T} \exp(\alpha \cdot \log M_{c\tau} + X'_{c\tau} \cdot \beta + \mu_t)} + \varepsilon_{ct}, \tag{33}$$

where S_{ct} is the number of drugs approved in category c at time t, divided by the total number of drugs approved in category c, and T is the total number of time periods in the sample. The advantage of this equation is that this transformation removes the drug category dummies, and therefore the coefficient of interest, α, can be estimated consistently, while at the same time taking out the fixed effects.

This still leaves the issue of how to estimate (33). We pursue two strategies. First, we estimate this equation using nonlinear least squares (NLLS). Second, we estimate it by maximum likelihood (ML) or quasi-maximum likelihood (QML). Woodridge (1999) shows that both NLLS and QML estimation strategies have good consistency properties, even when the true model is not Poisson. Finally, we will also report some estimates from the negative binomial model which relaxes the distributional of assumptions of the Poisson model in a maximum likelihood context (see, for example, Wooldridge, 2002, or Hausman, Hall, and Griliches, 1984).

3.2 Potential Market Size and Identification

There are at least two conceptual issues we need to discuss before estimating an equation similar to (28) or (30). First, there may be reverse causality in that new successful drugs may create markets for themselves, or at the very least, there may be other omitted characteristics influencing entry of new drugs, correlated with our measure of M_{ct}. As already noted, throughout we include drug category dummies and time dummies in (28), which control for fixed differences in innovation possibilities across drug categories, and time variation that might simultaneously affect the entry rates and the market size in the aggregate. More importantly, our market size measure M_{ct} is not the realized (ex post) market size of a drug, but the potentially-exogenous component of market size driven by demographic trends.

Second, as DiMasi et al (1991) report, from the time of initial research, it can take as long as 15 years for a drug to enter the market. In addition, changes in age demographics can be anticipated a long time in advance, so drug approvals may respond to anticipated future market sizes, as highlighted by our analysis in subsection 2.3. First, to reduce the impact of such delays on our estimates, we will consider ten-year intervals, as well as the shorter five-year intervals, for our units of observation. In addition, we will also experiment with leads and lags of $\log M_{ct}$ to determine whether there are significant delays and anticipation effects.

To implement our empirical strategy, we think of the categories as corresponding to drug

classes predominantly used by a specific age group. For example, Antibiotics such as Penicillin are used most by people aged 0-20. On the other hand, Antidepressants, like Prozac or Paxil, are used mainly by people aged 30-50 and older. We obtain the age composition of the users of various drugs from micro drug consumption data, and we combine this with the changes in U.S. demographics calculated from the CPS (Current Population Survey) data. Using these data sources, we construct our measure of potential market size as

$$M_{ct} = \sum_a u_{ca} \cdot p_{at}, \tag{34}$$

where p_{at} is the U.S. population at time t that is in age category a, and u_{ca} is the consumption (rate of use) of drug category c per individual in age group a. We compute M_{ct} in two alternative ways: first, using the U.S. population for p_{at} and the number of drugs used per person in age group a for u_{ca}; or using the total income of age group a for p_{at} and the expenditure per person in age group a for u_{ca}. The income-based measure corresponds more closely to market size in the theoretical model, which is a combination of the number of consumers and their incomes, and will be our main measure. For both measures, the over-time source of variation is not from patterns of use by various groups, but purely from demographic changes captured by p_{at}—u_{ca}'s are not time-varying. So for example, changes in prices, which potentially result from innovations and shift demand towards a particular type of drug, will not cause over-time variation in M_{ct}.

The most major threat to the validity of our empirical strategy is that there may be certain omitted non-market-size variables that affect the entry rate of new drugs in a potentially time-varying way (any variable that is not time-varying is taken out by the drug category fixed effects). If there are omitted variables related to market size or profit opportunities, this will lead to biased estimates, but it will not cause spurious positive estimates on our market size variable (in other words, the presence of such variables is essentially equivalent to mismeasurement of the appropriate market size, creating attenuation bias). We will also control for changes in drug coverage in health insurance policies as a potential source of omitted changes in demand. More threatening to our identification strategy would be omitted supply-side variables. If our instrument is valid, it should be orthogonal to variation in supply-side determinants of innovation. We will attempt to substantiate our identifying assumption further by adding a number of controls, and including lagged dependent variables, controls for pre-existing trends, and proxies for other incentives to undertake research in a particular field.[14]

[14] Another source of endogeneity may be that innovation in certain drug categories extend the lives of the elderly, and thus increasing their M_{ct}. Lichtenberg (2003) provides evidence that new drugs do indeed extend lives. This source of endogeneity is not likely to be quantitatively important, however, since the variation result-

4 Data and Descriptive Statistics

4.1 Basic Data Sources

We have gathered four types of data: demographic data, prescription drug classification schemes, prescription drug use, and FDA prescription drug approvals. The demographic data come from the March CPS, 1964-2000. We construct five age groups, 0-20, 20-30, 30-50, 50-60, and 60+. These divisions are motivated by the drug use patterns of these age groups. To construct income shares, we divide household income equally among the members of the household. Figure 2 shows population shares for the five age groups, and Figure 3 shows the corresponding income shares (to facilitate comparison with Figure 4, this figure starts in 1970). These figures show a large amount of variation across age groups over time. In particular, it is possible to trace the baby boomers, as the fraction of those in the age bracket 20-30 in the 1970s, and those in the age bracket 30-50 in the 1980s and the 1990s. These variations will be our main source of identification.

The FDA classifies all prescription drugs into 20 major drug categories, which are then further subdivided into 159 categories. These categories are based on a combination of therapeutic intent and chemical structure.[15] We drop 4 of the 20 major categories from this classification: Anesthetics, Antidotes, Radiopharmaceuticals and Miscellaneous.[16] With this procedure we obtain a total of 34 categories by breaking 10 of the 16 categories into finer groups when there were significant heterogeneity in terms of users' ages for the FDA categories. We separate the major categories of Antimicrobials, Psychopharmacologics, Nutrients, Hormones, Dermatologics, Neorologics, Ophthalmics, Otologics, Pain Relief and Respiratory because within these categories there were subcategories with significantly different age profiles of users. For example, within Antimicrobials, 0-20 year-olds use Antibiotics (except Tetracyclines) the most, while Antivirals are used most by people 30 and older. Appendix Table A1 lists the 34 categories.

Our main data source for drug use is the Medical Expenditure Panel Survey (MEPS), which is a sample of U.S. households over the years 1996-1998. The survey has age and income data for each household member, and covers about 25,000 individuals in each year. There is also

ing from extended lives in response to new drugs is a small fraction of the total variation in M_{ct}. Nevertheless, we will also report estimates that instrument M_{ct} by using past demographics, thus purging it from changing longevity.

[15] Other authors (e.g., Lichtenberg, 2003) have used a different classification system based on diseases. Since the FDA provides the therapeutic class for most drugs currently on the market, and some of our prescription drug use data is coded according to the FDA categories, we use this scheme.

[16] We drop Anesthetics, Radiopharmaceuticals and the Miscellaneous categories because most of the items in these categories were not developed for a distinct market. Radiopharmaceuticals are used for diagnostic purposes, and the Miscellaneous category is comprised mainly of surgical and dental tools. The Antidote category is dropped because there were few drugs approved and there was little use of the drugs in the surveys. See the Data Appendix for further details on the contruction of our categories.

a list of prescription drugs used by each person (if any), and the amount spent on drugs. In all, there are about 500,000 medications prescribed. We construct drug use and expenditure per person for each category by age. Appendix Table A1 contains the drug use per person for each age group for the 34 categories. This table shows a large amount of variation across drug categories. Many of the categories are used more by older people than by younger, but there are numerous exceptions. For example, Contraceptives are used most by 20-30 and 30-50 year-olds. On the other hand, Cardiovascular and Ophthalmic drugs are used primarily by individuals in the oldest category.

Combining the MEPS and the CPS, we construct our measures of potential market size according to equation (34). Our population-based measure calculates u_{ca} as the number of drugs per person used in category c by age group a from the MEPS and use total population in age group a at time t from the CPS for p_{at}. Our income-weighted measure uses expenditure per person on drug c for age group a from the MEPS, and total income of age group a at time t from the CPS for p_{at}.

We supplement the MEPS data with the National Ambulatory Medical Care Survey (NAMCS), which is an annual survey of doctors working in private practices. The survey includes drug use for the years 1980, 1981, 1985, and 1989-2000. Observations are at the doctor-patient-visit level; there are about 40,000 visits per year. Doctors are selected randomly, surveyed for a week, with patient-visits selected randomly from the corresponding week. The main use of NAMCS for us is that it covers a longer time period, enabling us to check whether the age composition of users across categories has changed over time. Using the NAMCS data, we construct a second drug categorization, with 30 categories. Appendix Table A2 compares this classification system with the 34 category system developed with the MEPS, listing which of the 159 FDA categories are in each of our categories. It should be noted that some of the 159 categories have been dropped from one classification system, but not the other, because there were not sufficient observations to construct reliable estimates of drug use from one of the surveys. As Appendix Table A2 shows, the two systems are closely related. In some cases, we have separated the FDA categories in one system but did not do this for the other (e.g., Pain Relief drugs). In general the 34 category system is a slightly less aggregated version of the other.[17]

Table 1 gives correlations between various measures of drug use. Panel A shows quite a high degree of correlation between the NAMCS surveys at various dates, indicating that the

[17] However, there are a number of cases where a given FDA category is combined with a second FDA category in one system, but with a third FDA category in the other system (e.g., Misc. Antibacterials are with Sulfonamides in the MEPS system, but with Antiseptics in the NAMCS system).

age profile of users has not changed significantly over the 1980s and the 1990s. The overall correlation row looks at the unconditional correlation. We also report weighted correlations, where observations are weighted by cell size in the drug use microdata (the MEPS or NAMCS). This is motivated by the fact that our estimates of the age distribution of users are more noisy when there are fewer individuals using drugs in a particular category. That the weighted correlation is uniformly greater than the overall correlation confirms this. The third row reports mean correlation by drug, which calculates the within category correlation between the two measures and then averages it across all categories. This measure is more informative for the question of whether the age distribution of the users for a particular drug has changed over time.

Panel B performs the same calculation for the three waves of the MEPS, and similarly shows a large amount of persistence in age distribution of users, though now the calculations refer to a shorter time frame.

Finally, panel C shows a high degree of correlation between expenditure per person and use per person in the MEPS data. But perhaps surprisingly, there is low correlation between the NAMCS and the MEPS. This is because the two surveys yield very different estimates for total use of each category (but very similar estimates of relative use by age groups within each category, as shown by the high level of mean correlation by drug). We conjecture that this reflects the fact that NAMCS, which samples doctors in private practice rather than individuals, is not as representative as the MEPS.

The last major data source is a list of FDA new drug approvals. This dataset contains the date at which the FDA approved the drug, as well as other information on the drug, such as which company submitted the drug for approval. We drop over-the-counter drugs and drugs that have the same identifying characteristics (i.e., same name, company, and category, or the same FDA approval number). This is because many FDA approvals (about 2,500 since 1970) are for a pre-existing drug with a new dosage level.[18] We focus on the time period 1970-2000. Since we can only match approvals for drugs that are still listed by the FDA, as we go back in time, the quality of the approvals data deteriorates.[19] In addition, because we are using age composition from the 1990s (or the 1980s and the 1990s for the NAMCS), as we go back in time, the quality of our measures of potential market size also deteriorates. Our approvals dataset for 1970-2000 includes 7,000 prescription drugs, which includes both generics and non-generics

[18] In addition, we also drop the so-called Orphan Drugs, which treat rare conditions, affecting fewer than 200,000 people. An example is botox, first developed to treat adult dystonia, which causes involuntary muscle contractions. We drop these drugs because we have difficulty matching them consistently, and also because they receive special inducements under the Orphan Drug Act.

[19] We use the earlier data when we estimate lagged depend variable models and control for pre-existing trends.

(see the Appendix for details). Since 1970, there have been about half as many approved non-generics as generics.

Figure 4 shows the log of drug approvals over time, to compare with changes in income shares depicted in Figure 3 (or population shares shown in Figure 2). To construct Figure 4, we compute drug approvals over five year intervals for the 34 categories. We then combine the 34 categories into five groups, based on the age group that most uses that category (thus this cut of the data uses only part of information that we will exploit in the regression analysis). Finally, we remove common year dummies from these series to make comparison easier, and plot the log approvals for each of the five groups.[20] Comparing this figure with Figure 3, a positive association between contemporaneous changes in population share and changes in drug approvals for the corresponding age group can be detected visually. For example, the population share of the 30-50 group increases steadily throughout the sample, and so does the entry of drugs most used by this group. Both the share of population and the entry of drugs for the age group 0-20, on the other hand, show a downward trend, while those for the 60+ age group show a mild increase followed by a mild decline. Table 2 also gives similar information about changes in (log) population by age group from the CPS and census data, and information about drug approvals for categories arranged according to the age group most using the drug in question. The information in Table 2 confirms the patterns in Figures 2-4. These patterns are explored in greater detail in the regression analysis below.

5 Results

5.1 Basic Specifications

Table 3A provides the basic results from the estimation of (32) and (31) with non-linear least-squares (NLLS) and ordinary least-squares (OLS), and Table 3B reports maximum likelihood estimates of the Poisson and negative binomial models. We start with the potential market size measure constructed using MEPS data. These basic specifications do not contain any covariates other than drug category effects and time effects. The results indicate large and significant effect of (potential) market size $\log M_{ct}$.

In column 1 of Table 3A, we start with our basic "income-weighted" measure of $\log M_{ct}$. constructed using expenditure data from the MEPS data set and income from the CPS, with

[20] There appear to be a number of institutional reasons behind the large fluctuations of approvals around a general upward trend. For example, in 1989, it was discovered that some FDA officials were taking bribes to speed up the approval process for generic drugs. As a result, in the early 1990's the approval process for generics was greatly slowed. See, for example, The Washington Post, August 16, 1989. When we separate our approval data into generics and non-generics, we see a large drop in generics approvals in the early 1990's, but only a small decline for non-generics. We thank Ernie Berndt for suggestions on this issue.

the time periods corresponding to five-year intervals. Observations are weighted by MEPS cell sizes, which is motivated by the fact that our estimates of age composition in smaller categories are substantially less precise, and unless otherwise stated, all standard errors are corrected for heteroscedasticity using the Huber-White formula. This basic specification leads to an estimate of α equal to 7.33 with standard error 1.81 using NLLS, which is significant at the 1 percent level. The OLS estimate is somewhat smaller, and less precisely estimated, 5.64, with standard error 2.52, thus significant only at the 5 percent level.[21] The quantitative magnitude of the effect in column 1 is plausible but large, implying that a 1 per cent increase in our market size measure leads to about a 6 to 7 percent increase in drug approvals.[22] In addition, the partial R^2 of the OLS specification is 0.18, suggesting that our measure of potential market size explains a sizable fraction of the total variation in the entry of new drugs.

Figure 5 shows a plot of the residuals of $\log N_{ct}$ against the residuals of $\log M_{ct}$ in the OLS regression, in both cases after drug category and period dummies are taken out. Observations are labeled by their drug category codes (see Appendix Table A1), and each code appears more than once, since there are multiple periods. The line in the figure corresponds to the estimated relationship reported in column 1. The figure shows that this relationship is not driven by outliers or some specific drug categories. Nevertheless, it can be seen that categories 42, 43 and 61 (Anorexiants, Central Nervous System drugs and Vitamins/Minerals) typically fit the pattern less well and are outliers in either direction. Motivated by this observation, we also estimated the basic regression using NLLS, and excluding these three categories. This leads to a very similar estimate of 7.50 with a standard error 2.72.

In column 2, we use ten-year intervals instead of the five-year intervals used in the previous columns. This is useful for two distinct reasons: first, there might be substantial noise in the entry of new drugs during the five-year interval; and second, given the lags and leads involved in the research process, ten-year intervals might better correspond to the relevant match between market size and entry (or innovation). The estimate of α using NLLS is now 5.58 with standard error 2.24, significant at the 1 percent level. While this is smaller than the estimate in column 1, the ten-year OLS estimate is slightly larger than the five-year estimate. In both cases the standard errors increase slightly in column 2. If five-year intervals were the appropriate time length, going to ten years should create attenuation bias, reducing the coefficient. The NLLS and OLS results are inconclusive on this point, and in the remainder, we look at both five-year

[21] If we cluster the standard errors at the level of drug categories, thus making an adjustment for potential serial correlation, the standard error increases from 2.52 to 2.71. Throughout, whether we do this type of clustering or not has little effect on standard errors. We also report lagged dependent variable specifications below as a direct way of dealing with potential serial correlation.

[22] We also ran this basic specification in levels (rather than logs), and this gives a similar, but less precise estimate, implying that to 1 percent increase in market size leads to a 3.4 percent increase in entry of new drugs.

and ten-year intervals.

Although the "income-weighted" measure of market size seems more satisfactory, in columns 3 and 4 we also look at the effects of changes in market size driven purely by population changes. Using this measure leads to estimates that are typically a little over 1 percentage point smaller, and are only slightly less precisely estimated.

Columns 1-4 weight the results by expenditure or number of users (cell size) of drugs in that category from the MEPS, since the age distribution of users in drug categories with few users is fairly noisy. To see the effect of weighting on the estimates, columns 5 and 6 report unweighted regressions. The results become less precise, and there is considerable attenuation. The NLLS estimates are now smaller than in columns 1 and 2, but still significant at the 1 percent, while OLS estimates are only significant at the 10 percent level. This indicates that there is a weaker relationship between market size and entry of new drugs within smaller drug categories. We believe that this reflects the substantially less precise estimation of age composition in those categories, which was also reflected by the correlations in Table 1. This conjecture receives further support from the fact that when we use age categories constructed from the NAMCS, which has a more even distribution of observations across drug categories, the unweighted results are also significant (see Table 7 below).

The results in Table 3B are broadly similar, and show that the main results are robust to different estimation methods. In panel A of that table, we use maximum likelihood. The number below the estimate is the maximum-likelihood standard error, while the number in curly brackets is the robust standard error that does not impose the Poisson structure to calculate the standard errors, and instead uses the Huber-White formula. The estimates are unweighted. In the first two columns, we compute market size using income and expenditure. The estimates are slightly larger than the corresponding estimates in columns 5 and 6 of Table 3A. In panel B we use a weighted maximum likelihood procedure, and obtain estimates that are similar to the estimates in columns 1-4 of Table 3A. Finally, we also report estimates from a negative binomial model in panel C, which allows for overdispersion of the Poisson parameter. In this case, our estimates are somewhat smaller than those in columns 1-4 of Table 3A, though still significant at the 5 percent level.

5.2 Delays and Anticipation Effects

The theoretical analysis suggested that delays in the development and approval processes are unlikely to create delays in the entry of new drugs in response to changes in market size, but there is room for new drugs to enter before the actual increase in market size because of antici-

pation effects, especially since demographic-driven changes in market size should be anticipated in advance. We investigate the role of delays and anticipation effects in this subsection by including lags ($\log M_{c,t-1}$) and leads ($\log M_{c,t+1}$) of potential market size on the right hand side of our estimating equation.

Column 1 of Table 4 replicates the basic specification from Table 3A for comparison. Columns 2 and 3 include the market size from the previous period, using five- and ten-year intervals. In column 2, the coefficient on current market size is larger than our baseline, 9.14 (standard error 3.03) with NLLS, and 11.10 (standard error 4.40) with OLS, and the coefficient on previous market size is negative (and statistically significant with NLLS, but not with OLS).[23] This pattern likely reflects the correlation between current and previous market size. Regardless, it is clear that new drug entry responds mainly to current market size. In column 3, the coefficients on current market size are smaller than the baseline and insignificant, but are much larger than the coefficients on past market size. Columns 4 and 5 report regressions that only have lagged market size. With five-year intervals, the coefficient is small and insignificant, and with ten-year intervals, it is negative and significant at the 1 percent level for NLLS. We conclude from these results that there is no evidence of significant delays in the response of entry of new drugs to changes in market size.

In column 6, we include the current and one period ahead market size in our estimation of the model. The estimate of current market size in panel A, 3.02, is considerably smaller than the baseline, and is insignificant. The coefficient on future market size is 8.14 and significant at the 1 percent level. Using ten-year intervals in column 7, both the current and future marke size are significant and approximately of the same magnitude. These results suggest that there is some anticipation effect, which is confirmed in columns 8 and 9, where we find a large and significant coefficient when we include only future market size. These results indicate that pharmaceutical companies respond to anticipated changes in demographics with five- or ten-years lags. This pattern is consistent with the fact that demographic trends are anticipated long in advance and with our theoretical results, which illustrated the possibility of limited anticipation effects.[24]

[23] We construct the lagged market size measures for 1960s using demographic information from the CPS, so the number of observations and does not decline.

[24] Here "limited" does not refer to the strength of the effect, but to the fact that the response to market size is 5-10 years before the change in market size, not further in advance. If we include further leads of market size, these are much smaller and insignificant. For example, the 15 years lead when included by itself is highly insignificant; the NLLS estimate is 3.63 with standard error equal to 4, and the OLS estimate is 4.29 with standard error of 5.34.

5.3 Changes in Health Insurance Coverage

Our market size measure only exploits changes in potential market size due to demographic trends. Another possible source of variation in market size comes from changes in coverage of drug expenditure in private or public health insurance programs. Finkelstein (2003), for example, exploits changes in the coverage of various vaccines to estimate the effect of these policies on the development of new vaccines.

During our sample period, there were significant changes in the coverage of drug expenditure in health insurance plans. For example, the percentage of 0-20 year-olds with some form of private health insurance coverage fell from about 73% to 69% between 1974 and 1996 (authors' calculations from the National Health Interview Survey). In the meantime, the corresponding percentage of 60+ year-olds rose from 62% to 75%. Furthermore, there have been changes in Medicaid eligibility rules, designed to insure more poor children. Since the age profile of drug use varies across categories, these changes in health insurance coverage induce additional changes in market sizes. We now investigate both whether controlling for this source of variation in potential market size affects the estimates of the impact of our measure M_{ct}, and whether insurance-induced changes in market size also have an effect on the entry of new drugs.

We use the National Health Interview Survey (NHIS, 1972-1996) to construct the fraction of each age group covered by a private health insurance plan. Because there is no consistent information on prescription drug coverage, we assign prescription coverage to any individual with both doctor and surgical coverage. Prescription drug coverage is highly correlated with this measure in the years where we can observe it. The NHIS also includes information on Medicaid, which covers prescription drugs, and on Medicare, which does not cover prescription drugs. The latter enables us to perform a "falsification test" to check the validity of our results.

Using the NHIS, with direct parallel to our M_{ct} measure, we construct the following variable:

$$H_{ct} = \sum_a u_{ca} \cdot p_{at} \cdot f_{at}, \tag{35}$$

where f_{at} is the fraction of age group a in period t with private health insurance, and u_{ca} and p_{at} are expenditure and income, as described above (in this section, we always use income based market size and weights). We add this variable to our estimating equations (31) and (32) as part of the covariate vector X_{ct}.[25] The estimation results are reported in Table 5.

The addition of $\log H_{ct}$ in column 1 has a small effect on the estimate of α relative to

[25] We can loosely think of $\log H_{ct}$ as the interaction between $\log M_{ct}$ and age-specific changes in insurance coverage rules over time, the f_{at}'s. Since the f_{at}'s do not very by drug category, we can think of their main effects being absorbed by the fixed effects. The analogy is only imperfect, however, since the interaction takes place when H_{ct} is constructed at the microlevel.

column 1 of Table 3A; the coefficient drops to 6.44 and the standard error increases to 3.11. The estimate of the coefficient of $\log H_{ct}$ is small and insignificant in both regression models. In column 2, we use ten-year intervals. The estimate of α decreases slightly (to 4.72 from the comparable estimate of 5.58 in Table 3A) and is no longer singificant at the 5 percent level.

In column 3, we use the fraction of people covered by Medicaid for f_{at}, with five-year intervals. Medicaid eligibility rules have changed to make children more likely to be covered. In the 1970s, about 9% of 0-20 year olds were covered by Medicaid; this fraction rose to 16% by the late 1990s. Exploiting these changes leads to an estimate of α nearly identical to the estimate in column 1. In column 7, we use the fraction covered by Medicare. Since Medicare does not cover prescription drugs, this should have no effect on the estimate of α, and the estimates there confirms this.

Columns 4, 5 and 6 repeat the same specifications in columns 1, 2 and 3 using the leads of the market size measure, $\log M_{ct}$ and confirm the findings in Table 4 of significant anticipation effects five to ten years before the actual changes in market size. In these specifications, we still include the current values of $\log H_{ct}$ (since there is less of an argument that there should be anticipation of future changes in this variable). The results are very similar if we instead include leads of this variable.

The results in columns 1-7 indicate no effect from changes in market size due to health insurance coverage, the $\log H_{ct}$ variables over and above $\log M_{ct}$, though their coefficients are always positive. This result might reflect the fact that $\log H_{ct}$ and $\log M_{ct}$ are positively correlated by construction, so estimating the effect of $\log H_{ct}$ separately is difficult. Whether $\log H_{ct}$ affects entry of new drugs is interesting in itself, and also informative for the question of whether the variable $\log M_{ct}$ is capturing the effect of market size or some other factor. Columns 8, 9 and 10 investigate this question by dropping $\log M_{ct}$. In column 8, we find that the effect of changes in market size, using private health insurance coverage, is significant at the 1 percent level using NLLS, and at the 10 percent level with OLS. The coefficient is somewhat smaller than the corresponding coefficient for $\log M_{ct}$, but broadly comparable. There are a number of reasons to expect the coefficients on the insurance-driven and demographic-driven market sizes to differ, most notably because the two affect different types of drug categories, and the demographic-driven changes are anticipated further in advance, potentially enabling a greater response. In column 9, we see that there is also a statistically significant effect of Medicaid insurance on the rate of entry of new drugs, but now the coefficient is much smaller.[26]

[26]Potential explanations include: (1) the fact that there is a negative correlation between Medicaid insurance and private insurance; (2) the fact that changes in Medicaid insurance affect only a few drug categories, specifically those for 0-20 year-olds, where the results may be less precise or the response of pharmaceutical firms less

In column 10, we repeat the same regression using Medicare. Since Medicare does not provide drug coverage, this should be interpreted as a "falsification exercise". It should be noted, however, that this is a fairly demanding falsification exercise, since $\log M_{ct}$ and $\log H_{ct}$ are correlated by construction. In any event, the results of this falsification exercise in column 10 is consistent with our presumption. Although the estimates are positive, they are not statistically significant, providing some support for our presumption that the positive effects of $\log M_{ct}$ and $\log H_{ct}$ we are estimating are not spurious.

5.4 Potential Supply-Side Determinants of Innovation

The main threat to our identification strategy is potential correlation between our market size measure and supply-side determinants of innovation, such as changes in scientific incentives, or in scientific opportunities as captured by the δ_j's in our model. Since such factors are difficult to observe, most of them will remain "omitted". In this subsection, we take a number of approaches to investigate the potential non-profit determinants of innovation.

First, it may be plausible to presume that changes in scientific constraints and opportunities may create changes in innovation rates. In terms of our model, the concern is that the δ_j's change over time (permanent differences in δ_j's are already taken out by our drug category fixed effects). If they do so, they may be serially correlated. In that case, we can partially control for the impact of changes in scientific opportunities by adding lags of $\log N_{ct}$ to our basic specifications.

Columns 1 and 2 of Table 6 report the results of estimating a lagged-dependent variable specification, by adding a one-period lag of the dependent variable, $\log N_{ct-1}$, to our basic specifications. In the OLS version, the basic regression therefore changes to:

$$\log N_{ct} = \alpha \cdot \log M_{ct} + \psi \cdot \log N_{ct-1} + \gamma \cdot d_{ct} + \delta_c + \phi_t + \varepsilon_{ct}. \tag{36}$$

Since $\log N_{ct-1}$ is correlated with the error term mechanically, OLS estimates to this equation would be biased, and we deal with this problem by instrumenting $\log N_{ct-1}$ with $\log N_{ct-2}$. This is a valid instrument as long as there is no serial correlation in the error term, ε_{ct} (see, for example, Blundell and Bond, 1998). This specification is also useful more generally if there are other sources of serial correlation, or some other reason for mean reversion.

pronounced; (3) the fact that the changes in market size resulting from Medicaid eligibility rules may be too small for drug companies to devote significant resources. To check the potential validity of the second hypothesis, we constructed a pseudo-market size measure equal to M_{ct} for the 0-20 age group, and 0 for older groups. Using this measure in our baseline specification of Table 3A, we obtain a coefficient of 0.34 with standard error 0.04, which is very similar to the estimate for Medicaid. This suggests that there is considerably less response to changes in demand coming from the youngest age group, and gives some support to the second hypothesis here.

The estimates of α from equation (36), reported in columns 1 and 2, are quite similar to the baseline. The lagged dependent variable, $\log N_{ct-1}$, is significant and negative for five-year intervals but not for ten-year intervals, suggesting that the degree of serial correlation due to scientific constraints or opportunities, or other sources, is relatively small.

Our second strategy looks at variation in the health benefits of new drugs across categories. New drugs in our data set include both drugs that are demanded by the consumers but do not "save lives", such as Prozac, Paxil, Vioxx, or Viagra, or those that actually save lives such as heart medicines or cancer treatments. A plausible conjecture is that non-profit incentives to develop drugs would be particularly responsive to opportunities to save lives or cure major illnesses. This conjecture receives support from the results of Lichtenberg (2003), which show the effect of pharmaceutical innovations on declines in mortality. To investigate this issue, we measure the number of life years lost corresponding to each drug category using the Mortality Detail Files from the National Center for Health Statistics from 1970-1994. Following Lichtenberg (2003), for each death, we subtract the person's age from 65, then calculate the sum of that number by drug and year. If someone dies at age 32, this counts as 33 life years lost; people dying older than 65 receive no weight in this calculation.[27]

We add this measure of life years lost to the right hand side of our baseline regression models as a proxy for this source of non-profit incentive to undertake research.[28] Since we are using mortality data prior to 1995, we drop the last time period from the regression. The baseline regression for the years 1970-1994 is reported in column 3 of Table 6, and is approximately the same size as the baseline using all years. Column 4 reports the result of using the life years lost variable, and finds no change in our estimate of α. The coefficient on the life years lost variable (unreported) is small and insignificant.[29] In the rest of this table, in addition to the NLLS estimates in panel A and the OLS estimates in panel B, which use current market size, we also report the NLLS estimate with leads of market size in panel C. In column 4, these lead results are similar to our baseline estimates in Table 4.

Third, we turn to another potential non-demand determinant of innovation, scientific funding. To investigate the implications of differences in scientific funding for various drug categories, we used the Computer Retrieval of Information on Scientific Projects (CRISP) dataset (details in the Data Appendix), and created a variable measuring total amount of federal fund-

[27] Note that the Mortality Detail Files are coded by disease class, so we must convert the classification to our system. Since many of our categories contain diseases or conditions that do not lead to death, we obtain many empty cells; typically there about 23 empty cells per five-year interval.

[28] We obtain very similar results if we take the log of this variable, and set log of 0 to 0.

[29] We also ran separate regressions using five- and ten-year lags of life years lost (both unreported), and again find no change in our estimates of α.

ing for research projects in all drug categories. We then include this variable as a control on the right hand side. To the extent that government funding also responds to potential market size (for example, because drug companies have a greater tendency to apply for funding in areas where they want to do research), this variable would be correlated with our market size measure. In practice, the correlation is low, and columns 5 and 6 show that the inclusion of this variable or both its current and lag values has little effect on our estimate of α

Fourth, to control for potential trends in scientific opportunities across drug categories, we add proxies for pre-existing trends. If pre-existing trends were important, we would expect the estimation of equation (36) to have shown an important role for lagged FDA approval. Nevertheless, we also look directly at the importance of pre-existing trends. To do this, we construct an estimate for pre-existing trends as: $\Delta_c = (\log N_{c,60} - \log N_{c,40})/2$, where $\log N_{c,40}$ is the log approvals for category c in 1940 and $\log N_{c,60}$ is the log approvals in 1960. We then estimate the equation:

$$\log N_{ct} = \alpha \cdot \log M_{ct} + \sum_{i=80,90} \Delta_c \cdot \sigma_i + \delta_c + \phi_t + \varepsilon_{ct}, \tag{37}$$

where σ_i's are period dummies for the 1980s and 1990s. This specification allows drug categories that have grown at different rates between 1940 and 1960 to also grow at different rates in the 1980s and the 1990s. Column 7 reports the results of this exercise. The coefficient on (log) market size is similar to our baseline estimate, 6.49, with standard error 2.32. Column 8 repeats the same exercise with $\Delta_c = (\log N_{c,70} - \log N_{c,40})/3$ as the measure of pre-existing trends. This estimate is also similar to our baseline. These results are perhaps not very surprising, since pre-1970 approvals are likely to be considerably noisier, thus only an imperfect control for pre-existing trends (in fact, the 1940-70 and 1970-2000 changes in approvals are essentially uncorrelated across drug categories).

An alternative strategy is to include linear time trends. This strategy is considerably more demanding, since using basically three ten-year observations, we now have to distinguish between linear trends and the effect of changes in market size, which are typically quite smooth. The estimating equation now becomes:

$$\log N_{ct} = \alpha \cdot \log M_{ct} + \delta_c + \phi_t + \eta_C \cdot t + \varepsilon_{ct},$$

where c refers to the 34 detailed drug categories, and C refers to the relevant 16 major drug category, i.e., the one which detailed category c belongs to. We expect technological differences to be captured by which of the 16 categories they belong to, since these categories were created based on therapeutic intent, while the subcategories are based on use by age group. The

estimates, reported in column 9, are smaller than our baseline, but still significant. For example, with NLLS, the estimate of α is 5.38 (s.e.=2.48). The OLS estimate is similar, but the larger standard error means it is insignificant. Using lead market size instead of current market size results in a similar pattern: the NLLS estimate, 10.65 (s.e.=3.92), is significant at the 1 percent level, while the (unreported) OLS estimate, 9.72 (s.e.=5.52) is only significant at 10 percent.

As a final check, we look at the potential effects of advances in biotechnology, such as the use of recombinant DNA, or other technological changes, during the late 1980s and the 1990s. In terms of our model, these developments would correspond to changes in the δ_j's. In column 10, we drop the categories of Cancer and Cardiovascular, which, according to the FDA approval list, have experienced the greatest number of Orphan Drugs. Although, as noted in footnote 18, our dependent variable does not include these drugs, it is useful to check that the results are robust to dropping the categories where entry of these drugs may have played an important role. The estimates in column 10 are nearly identical to the first column of Table 3A, showing that our results are not sensitive to dropping these two categories. In addition, there is anecdotal evidence that biotechnology firms were first active in producing insulin (the Glucose and Thyroid category) and in the Hematologic category.[30] In column 11, we drop these two categories, and again find that our results are essentially unchanged.

Finally, there is a group of drugs known as biologics, which include some vaccines, blood and plasma related products, and other products such as interferon and erythroproteins (used for red blood cell production). These drugs and products go through a separate FDA regulatory process, and are not included in our variable N_{ct}. Since biotechnology firms are particularly active in producing these types of drugs (especially, interferon and erythroproteins), we can assess the role of biotechnology firms by adding biologics approvals to our measure of drug approvals. The results of this regression, reported in column 12, show little change in the estimates of α.[31]

The results in this subsection therefore show that a number of controls for other (non-market-size related) determinants of the entry of new drugs have little effect on our main finding. These results are far from conclusive on the effect of scientific or other non-profit considerations in pharmaceutical research, since there may be other omitted characteristics.

[30] Biotechnology firms were also active in producing human growth factor, but we have not included this category in the analysis since there are only a small number of individuals using these drugs in the MEPS and NAMCS.

[31] We also dropped the categories where biologics are most common, Antivirals (12), Hematologics (20) and Immunologics (80), with little effect on our estimate of α.

Another possibility of raised by advances in biotechnology is that research may have become more "directed" in the late 1980s and the 1990s, for example, suggested by Galambos and Sturchio (1998), Gambardella (2000), and Malerba and Orsenigo (2000). We investigated this possibility by adding an interaction between $\log M_{ct}$ and a post-1990 dummy, and found no evidence for such a change.

Nevertheless, they are quite supportive of our identification strategy.

5.5 Reverse Causality

Lichtenberg (2003) shows that new drugs have increased the average age at death (and hence, to a lesser extent, life expectancy) by as much as 1 percent per year. This introduces the potential for reverse causality whereby the market size for successful drugs may be endogenously larger, because their users live longer. We think this is not a first-order concern, since the drug-induced changes in population are likely to be small relative to the large demographic changes that we are exploiting. Nevertheless, we address this issue by instrumenting for current population using the corresponding population from 10 years before. For example, we use the population fraction of 50+ year-olds in 1970 as an instrument for the population fraction of 60+ year-olds in 1980. The fraction of 50+ year-olds is highly correlated with the fraction of 60+ year-olds 10 years later, but is unaffected by new drugs that that are developed in the intervening 10 years, thus solving the reverse causality problem.

In column 2 of Table 7, we instrument for market size with past market size. With NLLS, the estimate is 6.90 (s.e. =2.95). This is slightly lower than the non-instrumented estimate for the same time period reported in column 1. Columns 3 and 4 show the corresponding results using ten-year intervals. The instrumented estimates in column 4 gives larger coefficients than the corresponding non-instrumented results. The NLLS estimate is significant at the 1 percent, while the linear model yields an estimates that is significant at the 10 percent. These results therefore show no evidence of reverse causality.

5.6 Results from the NAMCS

The rest of Table 7 repeats some of our main specifications using data from the NAMCS. This is a useful exercise for a number of reasons (despite the potential problems with the NAMCS noted above): first, the drug categories that we constructed are somewhat different between the MEPS and NAMCS; second, the NAMCS has a more even distribution of users across drug categories; and finally, because the NAMCS starts in 1980, we can check whether our use of drug consumption and expenditure data from the late 1990s introduces any biases.

Column 5 shows our baseline regression, with the same specification as column 3 of Table 3A (we cannot repeat the specifications of columns 1 and 2, since the NAMCS does not provide drug expenditure information). The estimate of α using NLLS is 3.60 with standard error 1.43, significant at the 5 percent level, which is considerably smaller than the corresponding MEPS estimate. Column 6 shows that this disparity is due to differences in samples, and not

due to differences in classification systems. In this column we use the MEPS data with the NAMCS classification system, and obtain estimates similar to column 3 of Table 3A. Therefore, the results differ most likely because we have different use per person estimates from the two surveys, which are likely due to the way the NAMCS samples are constructed.[32] Since the MEPS draws its sample from all U.S. households rather than from doctors in private practice, we believe that the MEPS gives better estimates for market size.

In column 7, we use ten-year time intervals. Both the NLLS and OLS are smaller than the five-year estimates in column 5, though broadly comparable. Columns 8 and 9 report the results without weighting by drug category size. In both columns, the estimates are somewhat larger than their counterparts in columns 5 and 7. This contrasts with the less precise unweighted estimates from the MEPS, and confirms our conjecture that the MEPS pattern was largely reflecting the less precise age distribution estimates for the smaller categories in that data set. The more even distribution of individuals across drug categories in NAMCS enables a more precise estimation of age distribution of users in the smaller cells.

Finally, we use the NAMCS to investigate whether our reliance on drug use data from the late 1990s induces any systematic bias. To see the potential reason for concern, suppose that a Gastrointestinal drug that is a major improvement over existing drugs enters the market before the first year of the MEPS, 1996, and is consequently used by a large number of 30-50 year-olds for the 1970-2000 period. The drug use and expenditure per person we estimate from the MEPS for Gastrointestinals would include the use and expenditure of that drug. As a result of the entry of this successful drug, we may overestimate the market size of Gastrointestinals for 30-50 year-olds. To correct for this possibility, we construct an alternative estimate of market size, M_{ct}^{1980} from NAMCS, utilizing the use per person numbers, u_{ca}^{1980}, only from the 1980 survey of NAMCS. We then estimate equation (33) post 1980, using NLLS and instrument $\log M_{ct}$ using $\log M_{ct}^{1980}$. For comparison, using our baseline measure only for 1980 onwards, we obtain an estimate of 6.04, with standard error 4.07 (column 10). The estimate using $\log M_{ct}^{1980}$ in column 11, on the other hand, leads to an estimate of α equal to 6.86 with standard error 4.15, significant at the 10 percent level, very similar to the estimate in column 10. This comparison suggests that using numbers from the MEPS does not lead to any significant bias.

[32] If we calculate market size using the fraction of total use for each age group rather than use per person for each age group, we obtain very similar estimates from the two surveys (which are closer to the MEPS estimates reported in the text). We consider this alternative measure of drug use to be less natural than expenditure per person, but it has the advantage of not being sensitive to the total expenditure by category.

5.7 Generics vs Non-Generics

Our analysis so far has combined generic and non-generic prescription drugs. As discussed in the theory section, although generic drugs may also correspond to a form of "innovation" in the sense that they divert business from incumbents and offer lower prices, the innovation process for developing a generic is different from that for a non-generic. It is therefore useful to separately look at the response of the entry of generics and non-generics to changes in market size.

Table 8 shows the results of some of our specifications using only generic drugs for N_{ct} with NLLS and OLS. Table 9 is similar and is for non-generics. Because we now look at generics and non-generics separately, there are more empty cells; for example, using all drugs and 5 year intervals, there are 7 empty cells, but there are 20 and 22 for generics and non-generics, respectively. Therefore, in Tables 8 and 9, we trust the NLLS results more than the OLS estimates, but we also report the latter for completeness.

The first two columns of both Tables 8 and 9 report the basic specifications for five- and ten-year intervals, similar to columns 1 and 2 in Table 3A. For generics, the NLLS estimate using the income-based measure of market size and five-year intervals (column 1) is 11.07 with standard error 2.29, significant at the 1 percent level. For comparison, the corresponding estimate in Table 9 for non-generic is 4.86, with standard error 2.07, smaller, though still significant at the 1 percent level. The larger estimate for generics suggests that the entry of generic drugs is more responsive to market size, but the results also show a significant response of non-generics to changes in market size.

In column 2 of both tables we use ten-year intervals, and find broadly similar results; there is a significant response of both generics and non-generics to market size, again with a stronger response from generics than non-generics.

In columns 3 and 4, we add lagged market size, and find no evidence of delayed responses for either generics or non-generics. In columns 5-8, we turn to anticipation effects. For generics, when both current and lead market sizes are entered together, the current market size is significant, and lead market size is insignificant. When lead market size is entered by itself, it is typically significant, but the standard errors are larger than the corresponding estimate losing current market size. This pattern suggests that the entry of generics is typically more responsive to current market size than to anticipated future market size. The pattern is somewhat different for non-generics. Here, when current and lead market sizes are present together, neither of them is significant, but lead market size typically has the right sign (i.e., positive) while current market size is zero or negative. When entered by itself, lead market size is not

significant at the 5 percent level, however; e.g., with NLLS, the coefficient on the five-years lead is 2.39 (s.e. = 2.78) and the ten-year lead is 2.32 (s.e. = 4.16). Therefore, there is some evidence that entry of non-generics is somewhat more responsive to anticipated future market size than current market size, but this evidence is weak.[33]

We can go one step further in our analysis of non-generic drugs. The FDA has labeled some generics as being "priority" drugs, meaning that they represent a significant therapeutic gain over existing drugs. The FDA has a second classification, whether or not a drug contains a new molecule. While many priority drugs are new molecules, a large fraction (about 30%) are not (and vice versa). Both types of drugs represent a significant amount of innovation, and potentially a different type of innovation compared to the the average non-generic. We combine them both in a new measure of N_{ct}. We use this smaller category in columns 9 and 10 of Table 9. In the basic specification, which is reported in column 9, we find an estimate of 3.59 with standard error 2.12, significant at the 10 percent level. This means that a 1 percent increase in market size leads to an approximately 4 percent increase in priority non-generics and new molecules. The relatively small estimate suggests that it may be more difficult to develop new molecular entities and priority drugs than other non-generics or generics, though this result may also reflect the larger standard errors resulting from the fact that there are many fewer new molecular entities than generics or non-generics in our data (a total of 400 drugs for 1970-2000).

6 Concluding Remarks

This paper argued that both theoretically and empirically, a key determinant of innovation is the potential market size of users. We provide evidence for this view from the pharmaceutical industry by exploiting changes in the potential market size for various drug categories driven by U.S. demographic changes. Our results indicate that a 1 percent increase in the potential market size for a drug category leads to approximately 4-7.5 percent growth in the entry of new drugs approved by the FDA. This is a substantial effect, and most of our various strategies find

[33] One reason to expect that non-generic entry may occur about a decade before the increase in market size would be the (limited) advantage of entering early highlighted in subsection 2.3. Generic entry, on the other hand, may be lagging behind non-generics due to the length of the patents. Since, in the U.S., patents are valid for about 17 to 20 years from the time they are granted, one might expect a longer lag than 5 or 10 years between non-generics and generics. Nevertheless, because drugs must go through a lengthy testing and approval process of about 10 years, pharmaceutical companies have considerably less the 17 to 20 years to market their drugs under patent protection. Consequently, a gap of about a decade between generic and non-generic entry may be plausible, though recall that the results here are not very conclusive.

Note also that in 1984, the Congress passed the Drug Price Competition and Patent Term Restoration Act. This Act allows patents to be extended for up to 5 years if the company could show that it lost marketing time while performing clinical trials or waiting for FDA approval. Under this Act, companies had a maximum of 14 years of patent protection under which to market their drugs.

it to be robust.

This finding, if further proven to be robust, has important implications both for research on the pharmaceutical industry, and for the endogenous growth and directed technical change literatures. It provides evidence that, as conjectured by these models, R&D and technological change are directed towards the most profitable areas. Furthermore, the magnitude of the effect is important for evaluating various theoretical predictions of these models. For example, directed technical change models suggest that the relative demand curves for factors can be upward, rather than downward, sloping if development of new technologies responds to a 1 percent increase in market size by more than 1 percent (see, for example, equations (21) and (22) in Acemoglu, 2002). Second, these findings imply that pharmaceutical research towards drugs with relatively small markets may be limited, which is a key premise of recent work by Kremer (2002). Building on this premise, Kremer suggests that there needs to be selective government incentives for developing drugs against malaria and other third-world diseases.

We view this research as part of a broader investigation of the effects of profit-incentives on innovation. Evidence from a single industry, in this case the pharmaceutical industry, may be nonrepresentative, especially because the pharmaceutical industry may be more research oriented than other industries. Future research both investigating the response of innovation and entry of new products in a specific industry, and also at the economy-wide level by comparing variation in the market size across different industries is necessary to substantiate the importance of market size and profit incentives on innovation.

7 References

Acemoglu, Daron. 1998. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics.* 113, pp. 1055-1090.

Acemoglu, Daron. 2002. "Directed Technical Change" *Review of Economic Studies.* 69, pp. 781-810.

Acemoglu, Daron. 2003. "Labor- and Capital-Augmenting Technical Change." *Journal of European Economic Associations*, Vol. 1, pp. 1-37.

Aghion, Philippe and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction" *Econometrica*, 60, pp. 323-351.

Aghion, Philippe and Peter Howitt. 1998. *Endogenous Growth Theory.* MA: MIT Press.

Ahmad, Syed. 1966. "On The Theory of Induced Invention," *Economic Journal.* 56, pp. 344-357.

Berndt, Ernst R., Iain M. Cockburn, Douglas L. Cocks, Arnold Epstein, M.D. and Zvi Grilliches. 1997. "Is Price Inflation Different for the Elderly? An Empirical Analysis of Prescription Drugs." *NBER Working Paper* #6182.

Blundell, Richard and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models" *Journal of Econometrics*, 87, pp. 115-144.

Ceruzzi, Paul E. 2000. *A History of Modern Computing.* MA: MIT Press.

Cockburn, Rebecca M. and Iain M. Henderson. 2001. "Scale and Scope in Drug Development: Unpacking the Advantages of Size in Pharmaceutical Research." *Journal of Health Economics*, 20, 1033-1057.

Danzon, Patricia M., Sean Nicholson and Nuno Sousa Pereira. 2003. "Productivity in Pharmaceutical-Biotechnology R&D: The Role of Experience in Alliances." *NBER Working Paper* #9615.

David, Paul. 1975. *Technical Choice, Innovation and Economic Growth: Essays on American and British Experience in the Nineteenth Century.* London:Cambridge University Press.

DiMasi, Joseph A., Ronald W. Hansen and Louis Lasagna. 1991. "Cost of Innovation in the Pharmaceutical Industry." *Journal of Health Economics*, 10, pp.107-142.

Drandakis, E. and Edmund Phelps. 1965. "A Model of Induced Invention, Growth and Distribution" *Economic Journal*, 76, pp. 823-840.

Finkelstein, Amy. 2002. "Health Policy and Technical Change: Evidence from the Vaccine Industry." mimeo.

Galambos, Louis and Jeffrey Sturchio. 1998. "Pharmaceutical Firms and the Transition to Biotechnology: A Study in Strategic Innovation." *Business History Review*, 72, pp. 250-278.

Gambardella, Alfonso. 2000. *Science and Innovation: The U.S. Pharmaceutical Industry During the 1980s*, Cambridge University Press, Cambridge UK.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, 25, pp. 501-522.

Griliches, Zvi and Jacob Schmookler. 1963. "Inventing and Maximizing" *American Economic Review*, 53, pp. 725-29.

Grossman, Gene and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy.* MA: MIT Press.

Hausman, Jerry, Bronwyn H. Hall and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application ot the Patents-R&D Relationship." *Econometrica*, 52, 909-938.

Hayami, Yujiro and Vernon Ruttan. 1970. "Factor Prices and Technical Change in Agricultural Development: the U.S. and Japan, 1880-1960" *Journal of Political Economy*, 77. pp. 1115-1141.

Henderson, Rebecca and Iain M. Cockburn. "Scale, Scope and Spillovers: The Determinants of Research Productivity in Drug Discovery." *Rand Journal of Economics,* 27, 32-59.

Kennedy, Charles 1964. "Induced Bias in Innovation and the Theory of Distribution" *Economic Journal,* 74, pp. 541-547.

Kling, Jim. 1998. "From Hypertension to Angina to Viagra." *Modern Drug Discovery,* 1 (2), pp. 31-8.

Kremer, Michael. 2002. "Pharmaceuticals and the Developing World" *Journal of Economic Perspectives*. 16 (4), pp. 67-90.

Lichtenberg, Frank R."The Allocation of Publicly Funded Biomedical Research," in *Medical Care Output and Productivity, Studies in Income and Wealth* volume 63. Ernst Berndt and David Cutler, eds. University of Chicago Press, 2001.

Lichtenberg, Frank R. 2002. "The Effects of Medicare on Health Care Utilization and Outcomes." *Frontiers in Health Policy Research*. 5, pp.

Lichtenberg, Frank R. 2002b. "Sources of U.S. Longevity Increases, 1960-1997." *NBER Working Paper* #8755.

Lichtenberg, Frank R. 2003. The Impact of New Drug Launches on Longevity: Evidence from Longitudinal, Disease-level Data from 52 Countries, 1982-2001. mimeo.

Lichtenberg, Frank R. and Suchin Virabhak 2002. "Pharmaceutical Embodied Technical Progress, Longevity and Quality of Life: Drugs as Equipment for Your Health." *NBER Working Paper* #9351.

Ling, Davina C., Ernie Berndt, and Richard G. Frank. 2003. "General Purpose Technolo-

gies, Capital-Skill Complementarity, and the Diffusion of New Psychotropic Medications among Medicaid Populations" MIT mimeo.

Malerba, Franco and Luigi Orsenigo. 2002. "Innovation and Market Structure in the Dynamics of the Pharmacetutical Industry: Towards a History Friendly Model." *Industrial and Corporate Change*. 11 (4), pp.667-703.

Mill, John Stuart. 1890. *Principles of Political Economy*, Oxford University Press.

Mokyr, Joel. 1990. *The Levers of Riches: Technological Creativity and Economic Progress.* NY:Oxford University Press.

Newell, Richard; Adam Jaffee and Robert Stavins. 1999. "The Induce Innovation Hypothesis and Energy-Saving Technological Change." *Quarterly Journal of Economics*. 114, pp. 907-940.

Nye, David E. 1990. *Electrifying America: Social Meanings of a New Technology, 1880-1940*. MA: The MIT Press.

Pakes, Ariel and Zvi Griliches. 1980. "Patents and R&D at the Firm Level: A First Look." *Economic Letters*, 5, 377-381.

Pakes, Ariel and Mark Schankerman. 1984. "An Explanation into the Determinants of Research Intensity" pp. 209-232 in *R&D, Patents and Productivity* edited by Zvi Griliches, National Bureau of Economic Research, University of Chicago Press, Chicago 1984.

Popp, David. 2002 "Induced Innovation and Energy Prices" *American Economic Review*. 92, pp. 160-180.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98. S71-S102.

Rosenberg, Nathan. 1974. "Science, Invention and Economic Growth" *Economic Journal*. 84, pp. 90-108.

Samuelson, Paul. 1965. "A Theory of Induced Innovations Along Kennedy-Weisacker Lines" *Review of Economics and Statistics*, 47, pp. 444-464.

Scherer, Michael. 1984. *Innovation and Growth*. MA: MIT press.

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

Scott Morton, Fiona. 1999. "Entry Decisions in the Generic Drug Industry." *The Rand Journal*, 30, pp. 421-440.

Wooldridge, Jeffrey M. 1999. "Distribution-free Estimation of Some Nonlinear Panel Data Models." *Journal of Econometrics*, 90, 77-97.

Wooldridge, Jeffrey M. 2002. *Econometrica Analysis of Cross Section and Panel Data*, MIT Press Cambridge.

8 Data Appendix

8.1 Prescription Drug Use and Expenditure, and Drug Categories from the MEPS

This section describes how we construct our measures of drug use and expenditure, as well as our classification system. The MEPS is an annual survey of randomly sampled households; we use the 1996, 1997 and 1998 surveys.

We obtain each person's age, the name of the prescription drug(s) used, and total expenditure (there are multiple records for people who use more than one prescription drug). Over the 3 years, we have about 500,000 drugs used and about 75,000 people. Expenditure includes out-of-pocket expenses, as well as amounts paid by insurers, which has been collected from surveys sent to insurance companies.

We begin with the 159 therapeutic categories, obtained from the FDA's National Drug Code (NDC) Directory. The names of these categories can be found in the second column of Appendix A1. The NDC Directory contains a file with the therapeutic category for most FDA approved drugs currently on the market. We assign each drug in the MEPS to one of the 159 categories by matching it by name with a drug in the NDC file. We cannot match about 10 percent of the drugs mentioned in the MEPS; these are usually not commonly used drugs, and make up less than 5 percent of the total drugs used.

We calculate drug use and expenditure by ten-year age group, using the survey weights. We divide these by the corresponding population and income for that year and age group, as estimated from the CPS, to obtain drug use and expenditure per person for each age group and category.

The FDA has assigned each of the 159 categories to one of 20 major therapeutic categories. Within each major category, we separate minor categories when there is sufficient heterogeneity in the age structure of drug expenditure (using drug use yields the identical classification system). From Table A1, it is apparent that we separate categories when there is considerable variation. For example, within Antimicrobials (categories 10-12) category 10 is used more by 0-20 year olds, category 11 has a steadily upward sloping profile, and category 12 is used approximately equally by individuals over age 30.

As noted in the text, we drop four major categories: Anesthetics, Antidotes, Radiopharmaceuticals, and Miscellaneous. We also drop several minor categories when there are not sufficient observations to estimate a reliable age structure. We use about 1,500 observations as our cutoff rule. We obtain this number from observing that only categories with more than 1,500 observations have fairly smooth age profiles.

Finally, wee aggregated the initial ten-year age groups into 5 age groups, 0-20, 20-30, 30-50, 50-60, 60+, by noting that most of our 34 categories show sharp peaks in one of the 5 groups.

8.2 Prescription Drug Use and Drug Categories from the NAMCS

The NAMCS differs from the MEPS in several important ways. First, it covers the years 1980, 1981, 1985, and 1989-2000. In constructing our classification system and estimating drug use, u_{ca}, we aggregate all years of NAMCS data.

Second, the survey is based on doctor-patient visits. It does not cover doctors at institutions or hospitals, unless they are considered to have a private practice. For each visit in the sample, there is a list of drugs prescribed, with a maximum of 5 to 8, depending on the year. Since we do not have information on expenditure, we weight multiple drugs for a single patient equally (as we do in constructing drug use per person with the MEPS). The survey also contains information on the doctor's primary diagnosis (which would be useful in weighting drug use), but it is not possible to create a consistent map between drugs prescribed and the diagnosis.

From 1993-2000, the NAMCS provides the FDA category for each prescribed medication.[34] We use this information to construct a mapping of medications to FDA class, which we can use to assign drugs from earlier survey years. Our worst success rate was in 1980, where we matched about 85% of prescribed drugs; in most years it was well above 90%. Because of these high rates, we believe that the bias from only using prescribed drugs in earlier years that were still being prescribed in the early 1990's is not large.

We initially construct drug use per person by ten-year age group as we do with the MEPS. We obtain 30 drug categories, as shown in Appendix Table A2. We find that the same 5 age groups are suitable for the NAMCS data.

We use the same cutoff rule for the NAMCS as for the MEPS for dropping FDA categories. It should be noted from Table A2 that we drop different FDA categories from the two surveys. For example, we drop Antifungals from the NAMCS system, and Anterior Pituitary from the MEPS.

8.3 Drug Approvals from the FDA

We have obtained a list of FDA drug approvals from Frank Lichtenberg. As noted in the text, this does not include biologics, which go through a separate licensing process. Also, we are only able to match a few orphan drugs, so we do not include them in the analysis. We keep only prescription drugs, because the NAMCS does not have information on over-the-counter drugs.

We match drugs in the approval list to FDA categories by drug name. We match 13,656 of 16,220 prescription drugs (84%) approved since 1970. Our success rate before 1970 is about 45%, which is why we restrict our main drug approvals data set to 1970-2000. We drop 4,451 of the matched drugs that have the same approval number and FDA class as a previously approved drug. We drop 1,103 drugs for which we have dropped the corresponding FDA category, because of insufficient observations in the MEPS (the corresponding number is 1,558

[34] Several drugs change FDA classes over the 8 years. In most cases, when we construct the 30 categories these drugs stay within the same category; we drop those that do not.

for the NAMCS). Finally, we drop drugs with the same name, MEPS category and different dosage from a previously approved drug, leaving us with our sample of 6,927 drugs (6,472 for the NAMCS). Of these, 2,089 are non-generics, and 429 are priority drugs and new molecules.

8.4 CRISP

The Computer Retrieval of Information on Scientific Projects (CRISP) dataset contains federally funded research projects at universities, hospitals and other institutions. Many of the grants are for very basic research, so they cover the very early stages of drug development. The projects are funded by the National Institutes of Health and the Substance Abuse and Mental Health Services Administration, as well as a variety of other agencies such as the FDA and the Center for Disease Control and Prevention. We have obtained a dataset with all pojects in the CRISP database for 1972-1995 from Frank Lichtenberg. In 1995 there were 57,553 grants, for a total of about 11 billion dollars.

Each record lists the project's investigator and affiliation, and the amount awarded. Most projects list one or several diseases which the researchers intend to study. For each disease listed, we know whether it is of primary, secondary or tertiary importance in the project. In our analysis, we use only primary diseases, and divide the award amount evenly between them. Our results are not sensitive to alternative weighting schemes.

The disease classification system is quite detailed; there are about 2,900 diseases, though these are arranged in a heirarchical structure into 35 major disease classes. We have mapped the detailed disease classes into our classification scheme. We are unable to map diseases into 5 of the smallest MEPS-based categories: Contraceptives, Skeletal Muscle Hyperactivity, Vertigo/Motion Sickness, Non-narcotic Analgesics, and Central Pain Syndromes. Our results are not sensitive to dropping these categories from the regressions.

Table 1:

Correlations Between Different Drug Use Measures

Panel A: NAMCS over time

	1980/1990	1990/2000	1980/2000
Overall Correlation	0.955	0.878	0.852
Weighted Correlation	0.967	0.862	0.862
Mean Correlation by Drug	0.832	0.794	0.741

Panel B: MEPS over time

	1996/1997	1997/1998	1996/1998
Overall Correlation	0.994	0.992	0.982
Weighted Correlation	0.998	0.997	0.993
Mean Correlation by Drug	0.851	0.888	0.854

Panel C: Correlation Between NAMCS and MEPS, and Between MEPS With and Without Income Weights

	MEPS/NAMCS	MEPS use/MEPS income weighted
Overall Correlation	0.180	0.969
Weighted Correlation	0.360	0.992
Mean Correlation by Drug	0.893	0.952

Notes: Overall correlation is the correlation of use per person (or expenditure per person for expenditure cells) across all categories. In weighted correlations, observations are weighted by cell size from the MEPS or NAMCS. Mean correlation by drug computes correlations separately by drug, then takes the average.

Table 2:

Population by Age Group and Drug Approvals Over Time

Panel A: Log Population by Age Group, From CPS

Age Group	1970-1980	1980-1990	1990-2000
0-20	18.13	18.07	18.15
20-30	17.34	17.52	17.44
30-50	17.69	17.94	18.20
50-60	16.91	16.92	17.02
60+	17.21	17.43	17.54

Panel B: Log Drug Approvals of Generics by Age Group, From FDA

Age Group	1970-1980	1980-1990	1990-2000
0-20	6.43	5.77	4.30
20-30	4.08	4.14	3.14
30-50	6.13	6.82	6.13
50-60	5.57	5.76	5.36
60+	6.33	7.00	6.33

Panel C: Log Drug Approvals of Non-generics by Age Group, From FDA

Age Group	1970-1980	1980-1990	1990-2000
0-20	5.54	5.89	5.39
20-30	3.74	4.13	3.69
30-50	5.08	5.48	6.04
50-60	4.74	4.98	5.48
60+	5.53	6.58	6.28

Notes: In Panel A, log population by age group is the log of mean population over the ten year interval, using the March CPS. In panels B and C, each of the 34 drug categories is assigned one age category, based on the age group that uses that category most. Log of drug approvals is the log of the sum of all approvals in the indicated 10 year interval, for all drug categories in the given age category.

Table 3A:

Effect of Changes in Market Size on New Drug Approvals

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: NLLS for Poisson model, dependent variable is count of drug approvals					
Log Market Size	7.33 (1.81)	5.58 (2.24)	6.12 (1.71)	4.31 (1.92)	5.41 (1.71)	4.51 (1.89)
R Squared	0.88	0.92	0.88	0.93	0.78	0.86
	Panel B: OLS, dependent variable is log drug approvals					
Log Market Size	5.64 (2.52)	5.97 (2.85)	5.05 (2.19)	5.09 (1.94)	3.28 (1.80)	4.39 (2.53)
R Squared	0.87	0.92	0.87	0.92	0.82	0.88
Number of Observations	204	102	204	102	204	102
Length of Time Interval (Years)	5	10	5	10	5	10
Drug Category Weights	Yes	Yes	Yes	Yes	No	No
Market Size and Weights Include Income	Yes	Yes	No	No	Yes	Yes

Notes: Counts of drug approvals are computed from the FDA dataset of New Drug Approvals, by counting drug approvals for each category over five- and ten-year intervals (see Appendix for details). Market Size is obtained by multiplying the time-invariant expenditure per person of users in a particular age group, calculated from the MEPS, by total income of that age group at that date, from the CPS, and summing over all age groups. When market size does not include income, use per person is multiplied by population. See text for details. Huber-White robust standard errors are reported in parentheses. All regressions include drug and period dummies, and the 34 drug categories constructed from the MEPS, as described in the Appendix. In panel A, the Poisson model is estimated by NLLS (with the Hausman, Hall and Griliches, 1984, transformation). In panel B, if a cell is empty, log approval is set equal to zero, and a dummy variable, equal to 1 when the cell is empty, is added to the regression. Regression weights are cell size for the category from the MEPS, either total expenditure or total use.

Table 3B:

Effect of Changes in Market Size on New Drug Approvals

	(1)	(2)	(3)	(4)
	Panel A: Poisson ML, dependent variable is count of drug approvals			
Log Market Size	6.41 (0.68) {2.92}	6.09 (0.73) {2.02}	5.78 (0.65) {1.75}	5.16 (0.70) {1.59}
	Panel B: Weighted Poisson ML, dependent variable is count of drug approvals			
Log Market Size	7.36 {2.56}	6.79 {2.01}	7.35 {1.45}	5.95 {1.15}
	Panel C: Negative Binomial ML, dependent variable is count of drug approvals			
Log Market Size	4.74 (1.77) {2.27}	4.80 (1.89) {2.11}	4.42 (1.89) {2.51}	4.55 (1.96) {2.23}
Number of Observations	204	102	204	102
Length of Time Interval (Years)	5	10	5	10
Market Size and Weights Include Income	Yes	Yes	No	No

Notes: Drug approvals and market size variables, and regression weights are constructed as in Table 3A. All regressions include drug and period dummies, and use the 34 MEPS-based drug categories. In panels A and B a Poisson model is estimated using maximum likelihood. In panel C a negative binomial model is estimated using maximum likelihood. Maximum Likelihood standard errors in parentheses, and Huber-White standard errors in curly brackets.

Table 4:

Delays and Anticipation Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Panel A: NLLS, dependent variable is count of drug approvals					
Log Market Size	7.33 (1.81)	9.14 (3.03)	4.03 (2.56)			3.02 (3.35)	6.94 (2.29)		
Lag Market Size		-3.98 (1.61)	-2.30 (2.21)	0.71 (0.43)	-4.12 (2.00)				
Lead Market Size						8.14 (3.75)	8.57 (3.36)	11.37 (2.36)	11.77 (3.42)
R Squared	0.88	0.87	0.92	0.87	0.92	0.63	0.78	0.60	0.75
				Panel B: OLS, dependent variable is log drug approvals					
Log Market Size	5.64 (2.52)	11.10 (4.40)	5.21 (3.45)			3.69 (3.26)	7.24 (3.57)		
Lag Market Size		-4.95 (3.50)	-1.78 (2.85)	2.30 (1.78)	-3.90 (2.20)				
Lead Market Size						3.09 (4.07)	9.59 (6.67)	5.74 (3.10)	12.46 (7.30)
R Squared	0.87	0.88	0.92	0.87	0.92	0.78	0.96	0.90	0.96
Number of Observations	204	204	102	204	102	170	68	170	68
Length of Time Interval (Years)	5	5	10	5	10	5	10	5	10

Notes: Drug approvals and market size variables are constructed as in Table 3A. Lag Market Size refers to one-period lag of Log Market Size, and Lead Market Size refers to one-period lead of Log Market Size. All regressions use income-based market size and income-based weights. Regressions include drug and period dummies and all 34 drug categories. In panel A, the Poisson model is estimated using NLLS (with the Hausman, Hall and Griliches, 1984, transformation), and in panel B empty approval cells are set equal to zero and a dummy variable for empty cells is added to the OLS regression, as in Table 3A. Huber-White standard errors in parentheses.

Table 5:

Controlling for Changes in Health Insurance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				Panel A: NLLS for Poisson model, dependent variable is count of drug approvals						
Log Market Size	6.44 (3.11)	4.72 (3.65)	7.12 (1.84)				7.66 (1.82)			
Lead Market Size				9.57 (2.84)	10.46 (6.51)	10.49 (2.45)				
Log Market Size of Insured	1.07 (2.13)	0.84 (2.64)	0.11 (0.12)	1.72 (1.55)	0.48 (1.82)	0.12 (0.12)	0.74 (2.05)	4.68 (1.29)	0.21 (0.12)	1.59 (2.36)
R Squared	0.88	0.92	0.88	0.60	0.56	0.60	0.88	0.88	0.87	0.86
				Panel B: OLS, dependent variable is log drug approvals						
Log Market Size	5.85 (3.96)	6.59 (5.15)	5.09 (2.57)				5.77 (2.53)			
Lead Market Size				8.60 (3.36)	12.46 (7.51)	7.44 (3.14)				
Log Market Size of Insured	-0.05 (2.31)	-0.55 (2.89)	0.17 (0.06)	0.19 (1.38)	2.48 (1.82)	0.20 (0.06)	0.89 (1.84)	2.72 (1.43)	0.21 (0.06)	0.94 (2.06)
R Squared	0.87	0.92	0.88	0.90	0.96	0.90	0.87	0.87	0.87	0.87
Number of Observations	204	204	102	170	68	170	204	204	204	204
Length of Time Interval (Years)	5	5	10	5	10	5	5	5	5	5
Type of Insurance	Any Private	Any Private	Medicaid	Any Private	Any Private	Medicaid	Medicare	Any Private	Medicaid	Medicare

Notes: Drug approvals and market size variables are constructed as in Table 3A. Market size of insured is obtained by multiplying the time-invariant expenditure per person of users in a particular age group, by total income of that age group, by the fraction of people in that age group with the corresponding type of health insurance, as calculated from the NHIS, and summing over all age groups. See text for details. All regressions use income-based market size and income-based weights. Regressions include drug and period dummies and all 34 drug categories. In panel A, the Possion model is estimated using NLLS (with the Hausman, Hall and Griliches, 1984, transformation), and in panel B empty approval cells are set equal to zero and a dummy variable for empty cells is added to the OLS regression, as in Table 3A. Huber-White standard errors in parentheses.

Table 6:

Potential Supply-Side Determinants of Innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
			Panel A: NLLS for Poisson model, dependent variable is count of drug approvals									
Log Market Size	8.28 (1.88)	7.80 (2.04)	8.21 (1.95)	7.93 (2.01)	7.17 (1.87)	7.19 (1.89)	6.49 (2.32)	7.10 (2.20)	5.38 (2.48)	7.03 (1.83)	7.17 (1.86)	7.65 (1.92)
Lagged Dependent Variable	-1.75 (0.62)	0.52 (0.46)										
R Squared	0.88	0.94	0.89	0.89	0.87	0.87	0.92	0.93	0.92	0.87	0.87	0.87
			Panel B: Linear Regressions, dependent variable is log drug approvals									
Log Market Size	6.01 (2.59)	7.33 (3.30)	7.16 (2.70)	6.79 (2.74)	5.69 (2.41)	5.48 (2.73)	6.18 (2.76)	6.00 (3.13)	5.07 (3.63)	5.74 (2.63)	5.69 (2.51)	5.85 (2.59)
Lagged Dependent Variable	-1.48 (0.83)	0.53 (0.94)										
R Squared	0.88	0.93	0.90	0.89	0.88	0.89	0.92	0.93	0.95	0.78	0.88	0.87
			Panel C: NLLS for Poisson model, dependent variable is count of drug approvals									
Lead Market Size			11.37 (2.36)	11.39 (2.39)	11.74 (2.37)	11.48 (2.34)	12.11 (3.50)	13.56 (3.39)	10.65 (3.92)	10.32 (4.01)	14.03 (4.43)	9.39 (4.09)
R Squared			0.60	0.60	0.60	0.62	0.75	0.78	0.97	0.59	0.51	0.60
Length of Time Interval (Years)	5	10	5	5	5	5	10	10	10	5	5	5
Life Years Lost	No	No	No	Yes	No	No	No	No	No	No	No	No
CRISP Funding	No	No	No	No	Yes	Yes	No	No	No	No	No	No
Pre-existing Trends Interacted with Period Dummies	No	No	No	No	No	No	Yes	Yes	No	No	No	No
Major Drug Category	No	No	No	No	No	No	No	No	Yes	No	No	No
Categories Excluded	None	None	None	None	None	None	None	None	None	Cancer, Cardio	Insluin, Thyroid	Biologics Included

Notes to Table 6: Drug approvals and market size variables are constructed as described in Table 3A. In all regressions, market size is computed using income and expenditure shares. Huber-White standard errors in parentheses. All regressions include 34 categories, drug and period dummies, and are weighted by cell size. In panels A and B, the number of observation is 204 in columns 1, 5, and 12; 102 in 2, 7, 8, and 9, 170 in 3, 4 and 6; and 192 in 10 and 11. In panel C, the number of observation is 170 in columns 3, 4, 5, 11, and 12; 68 in 7, 8, and 9; and 160 in 10 and 11. In columns 1 and 2 the lagged dependent variable is instrumented with the twice lagged dependent variable. Life years lost is defined as the number of years prior to age 65 for each death in the US, as calculated from the Mortality Detail Files. See text for details. Column 4 includes a count of total life years lost due to diseases in the corresponding category and time interval. Columns 5 and 6 include the amount of funding from NIH grants for research in each category in the particular interval, as calculated from the CRISP database (see Appendix for details). Column 6 also includes the lag of this variable. 1940/1960 trend for category c is the log difference of drug approvals for category c between 1960 and 1940. In column 7, the 1940/1960 trend is interacted with period dummies for the 1980's and 1990's decades. Column 8 reports the corresponding regressions for the 1940/1970 trends. See text for details. The interactions were generally insignificant, and are not reported. Major drug category trends are linear time trends interacted with dummies to which of the 16 major drug categories that each category belongs. See text for details. In column 12 FDA approvals of biologics for each category and time interval are added to the dependent variable.

Table 7:

Instrumenting for Market Size and Results Using NAMCS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					Panel A: NLLS, dependent variable is count of drug approvals						
Log Market Size	7.33 (1.81)	6.90 (2.05)	5.58 (2.24)	7.79 (2.89)	3.60 (1.43)	5.28 (2.12)	3.07 (1.80)	4.86 (1.91)	4.25 (2.29)	6.04 (4.07)	6.86 (4.15)
R Squared	0.88	0.88	0.92	0.92	0.79	0.86	0.78	0.71	0.72	0.87	0.86
					Panel B: Linear Regressions, dependent variable is log of drug approvals						
Log Market Size	5.64 (2.52)	5.35 (2.80)	5.97 (2.85)	7.46 (4.02)	4.11 (1.54)	6.37 (2.64)	4.69 (2.00)	5.11 (2.03)	6.86 (3.27)	7.66 (3.77)	8.80 (4.38)
R Squared	0.87	0.87	0.92	0.92	0.87	0.86	0.93	0.83	0.89	0.90	0.90
Number of Observations	204	204	102	102	180	180	90	180	90	120	120
Number of Categories	34	34	34	34	30	30	30	30	30	30	30
Length of Time Interval (Years)	5	5	10	10	5	5	10	5	10	5	5
Instrument for Market Size with Previous Market Size	No	Yes	No	Yes	No	No	No	No	No	No	No
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Instrument Market Size with 1980 Based Market Size	No	No	No	No	No	No	No	No	No	No	Yes

Notes: In columns 1-4 and 6, market size is computed as in Table 3A, and is income-based for columns 1-4. In columns 5 and 7-11, market size is computed as in Table 3A, except using the NAMCS instead of the MEPS. Huber-White standard errors in parentheses. In columns 5-11, regressions include 30 drug categories, constructed from the NAMCS and in the same manner as the MEPS, as explained in the Appendix. All regressions include drug and period dummies. Regression weights in columns 1-4 are total expenditure of the category, as computed from the MEPS. Regression weights in columns 5 and 7-11 are total use for the category, computed from the NAMCS. Regression weights in column 6 are total use, computed from the MEPS. In columns 1 and 2, current market size is instrumented with the market size 10 years earlier of the age group that is 10 years younger. For example, the market size of 20-30 year-olds in 1970 is instrumented by the market size of 10-20 year-olds in 1960.1980-based market size is constructed in the same way as market size, except that only the 1980 NAMCS data is used. In column 9 current market size is instrumented with this variable.

Table 8:

Generics: Effect of Changes in Market Size on New Drug Approvals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: NLLS, dependent variable is count of drug approvals				
Log Market Size	11.07 (2.29)	9.04 (2.59)	10.76 (3.49)	9.00 (3.24)	8.96 (4.30)	6.99 (3.55)		
Lag Market Size			2.17 (2.73)	0.65 (2.55)				
Lead Market Size					2.15 (4.70)	4.04 4.92	11.20 (2.85)	7.20 (4.79)
R Squared	0.83	0.90	0.90	0.90	0.57	0.87	0.56	0.86
				Panel B: OLS, dependent variable is log drug approvals				
Log Market Size	10.49 (3.79)	10.96 (4.19)	12.02 (4.90)	10.95 (4.78)	9.34 (5.13)	9.41 (4.06)		
Lag Market Size			2.43 (3.86)	1.88 (3.69)				
Lead Market Size					1.35 (6.40)	6.05 (10.44)	10.09 (4.21)	10.13 (11.03)
R Squared	0.83	0.92	0.92		0.86	0.96		0.95
Number of Observations	204	102	204	102	170	68	170	68
Length of Time Interval (Years)	5	10	5	10	5	10	5	10

Notes: Dependent variables include only approvals of generic drugs. Market size is the income-based measure, constructed as in Table 3A. All regressions include period and category dummies, and are weighted by category size as described in Table 3A. Hubler-White standardard errors in parentheses.

Table 9:

Non - Generics: Effect of Changes in Market Size on New Drug Approvals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Panel A: NLLS, dependent variable is count of drug approvals					
Log Market Size	4.86 (2.07)	5.18 (2.45)	8.75 (3.00)	5.18 (2.59)	-2.63 (4.03)	2.06 (5.34)			3.59 (2.12)	3.40 (2.67)
Lag Market Size			-6.54 (3.04)	0.02 (2.60)						
Lead Market Size					4.62 (4.59)	-4.17 (7.12)	2.39 (2.78)	2.32 (4.16)		
R Squared	0.82	0.89	0.82	0.89	0.23	0.35	0.24	0.33	0.82	0.90
					Panel B: OLS, dependent variable is log drug approvals					
Log Market Size	3.19 (2.53)	4.81 (2.98)	7.68 (5.74)	5.19 (3.53)	-0.27 (4.45)	5.97 (4.14)			1.63 (2.50)	0.28 (3.80)
Lag Market Size			-5.27 (3.74)	0.90 (3.65)						
Lead Market Size					2.35 (6.26)	-7.47 (7.76)	1.82 (3.54)	-5.03 (6.77)		
R Squared	0.80	0.87	0.82	0.87	0.85	0.94	0.88	0.94	0.83	0.84
Number of Observations	204	102	204	102	170	68	170	68	204	102
Length of Time Interval (Years)	5	10	5	10	5	10	5	10	5	10

Notes: Dependent variables in columns 1-8 include only approvals of non-generic drugs. Dependent variables in columns 9 and 10 include only priority non-generics and new mocecules, as described in the Appendix. Market size is the income-based measure, constructed as in Table 3A. All regressions include period and category dummies, and are weighted by category size as described in Table 3A. Huber-White standard errors in parentheses.

Appendix Table A1:

Summary of Classifications and Drug Use by Age Group, Computed From MEPS

Class	Description	Share of Use 0-20	20-30	30-50	50-60	60+
10	Penicillins, Cephalosporins, Lincosamides, Sulfonamides, Misc. Antibacterials	0.61	0.30	0.38	0.44	0.45
11	Tetracyclines, Urinary Tract Antiseptics, Quinolones	0.02	0.06	0.06	0.09	0.12
12	Antifungals, Antivirals	0.03	0.03	0.08	0.09	0.07
20	Hematologics	0.00	0.00	0.03	0.11	0.43
30	Cardiovascular - Renal	0.05	0.10	0.69	2.68	6.05
40	Sedatives/Hypnotics, Antianxiety	0.01	0.04	0.16	0.27	0.41
41	Antipsychotics/Antimania, Antidepressants	0.08	0.19	0.57	0.70	0.57
42	Anorexiants	0.04	0.01	0.03	0.02	0.01
43	Misc. Central Nervous System	0.11	0.01	0.01	0.01	0.02
50	Gastrointestinals	0.03	0.08	0.24	0.52	0.83
60	Hyperlipidemia, Electrolyte Replenishment/Regulation, Calcium Metabolism	0.01	0.01	0.13	0.67	1.37
61	Vitamins/Minerals	0.06	0.09	0.07	0.09	0.13
70	Adrenal Corticosteroids	0.05	0.04	0.09	0.14	0.24
71	Androgens/Anabolic Steroids, Estrogens/Progestins	0.02	0.38	0.34	1.30	0.67
72	Blood Glucose, Thyroid	0.02	0.10	0.35	1.02	1.70
73	Contraceptives	0.01	0.21	0.10	0.01	0.01
80	Immunologics	0.00	0.01	0.02	0.02	0.03

Appedix Table A1 (cont.)

		Share of Use				
Class	Description	0-20	20-30	30-50	50-60	60+
90	Dermatologics, Topical Anti-Infectives	0.09	0.09	0.09	0.12	0.17
91	Topical Steroids	0.01	0.01	0.01	0.02	0.04
100	Extrapyramidal Movement	0.00	0.01	0.03	0.03	0.08
101	Skeletal Muscle Hyperactivity, Anticonvulsants	0.05	0.11	0.26	0.29	0.27
110	Oncolytics	0.00	0.01	0.05	0.16	0.17
120	Misc. Ophthalmics, Glaucoma	0.00	0.00	0.01	0.06	0.41
121	Ocular Anti-Infective	0.06	0.05	0.05	0.08	0.16
130	Topical Otics	0.02	0.01	0.01	0.03	0.04
131	Vertigo/Motion Sickness	0.02	0.01	0.03	0.05	0.13
140	General Analgesics, Narcotic Analgesics, Antiarthritics, NSAID	0.09	0.29	0.59	0.89	1.13
141	Non-Narcotic Analgesics	0.00	0.01	0.02	0.04	0.07
142	Antigout	0.00	0.00	0.01	0.06	0.14
143	Central Pain Syndromes	0.01	0.01	0.02	0.02	0.01
150	Antiparasitics	0.00	0.01	0.03	0.03	0.05
160	Antiasthmatics, Nasal Decongestants	0.20	0.14	0.22	0.47	0.66
161	Antitussives, Antihistamines, Corticosteroids	0.15	0.20	0.29	0.45	0.41
162	Cold Remedies	0.07	0.05	0.07	0.08	0.08

Appendix Table A2:

Comparison of MEPS and NAMCS Classification Systems

MEPS	FDA Categories Included	NAMCS	FDA Categories Included
10	Penicillins, Cephalosporins, Lincosamides, Sulfonamides, Misc. Antibacterials	10	Penicillins, Cephalosporins, Lincosamides, Aminoglycosides, Sulfonamides
11	Tetracyclines, Urinary Tract Antiseptics, Quinolones	11	Tetracyclines
12	Antifungals, Antivirals	12	Urinary Tract Antispetics, Misc. Antibacterials, Quinolones
		13	Antivirals
20	Hematologics	20	Deficiency Anemias
		21	Anticoagulants
30	Cardiovascular - Renal	30	Cardiovascular - Renal
40	Sedatives/Hypnotics, Antianxiety	40	Sedatives/Hypnotics, Antianxiety
41	Antipsychotics/Antimania, Antidepressants	41	Antipsychotics/Antimania, Antidepressants
42	Anorexiants	42	Anorexiants
43 50	Misc. Central Nervous System Gastrointestinals	50	Acid/Peptic Disorders, Laxatives
		51	Antidiarrheals, Antispasmodics/Anticholinergics
60	Hyperlipidemia, Electrolyte Replenishment/Regulation, Calcium Metabolism	60	Hyperlipidemia, Electrolyte Replenishment/Regulation
61	Vitamins/Minerals	61	Vitamins/Minerals
70	Adrenal Corticosteroids	70	Adrenal Corticosteroids, Anterior Pituitary
71	Androgens/Anabolic Steroids, Estrogens/Progestins	71	Androgens/Anabolic Steroids, Blood Glucose, Thyroid
72	Blood Glucose, Thyroid	72	Estrogens/Progestins
MEPS	FDA Categories Included	NAMCS	FDA Categories Included
73	Contraceptives	73	Contraceptives, Infertility, Uterine Relaxants/Stimulants
80	Immunologics	80	Allergenic Extracts
		81	Immune Serums, Vaccines

Appendix Table A2 (cont.)

MEPS	FDA Categories Included	NAMCS	FDA Categories Included
90	Dermatologics, Topical Anti-Infectives	90	Skin/Mucous Membranes
91	Topical Steroids		
100	Extrapyramidal Movement	100	Extrapyramidal Movement, Myasthenia Gravis
101	Skeletal Muscle Hyperactivity, Anticonvulsants	101	Skeletal Muscle Hyperactivity, Anticonvulsants
110	Oncolytics	110	Oncolytics
120	Misc. Ophthalmics, Glaucoma	120	Ophthalmics
121	Ocular Anti-Infective		
130	Topical Otics	130	Otics
131	Vertigo/Motion Sickness		
140	General Analgesics, Narcotic Analgesics, Antiarthritics, NSAID	140	Pain Relief
141	Non-Narcotic Analgesics		
142	Antigout		
143	Central Pain Syndromes		
150	Antiparasitics	150	Antiparisitics
160	Antiasthmatics, Nasal Decongestants	160	Antiasthmatics/ Bronchodilators
161	Antitussives, Antihistamines, Corticosteroids	161	Nasal Decongestans, Antitussives, Antihistamines, Cold Remedies, Lozenges, Corticosteroids
162	Cold Remedies		

Figure 1: Response to an Anticipated Increase in Market Size

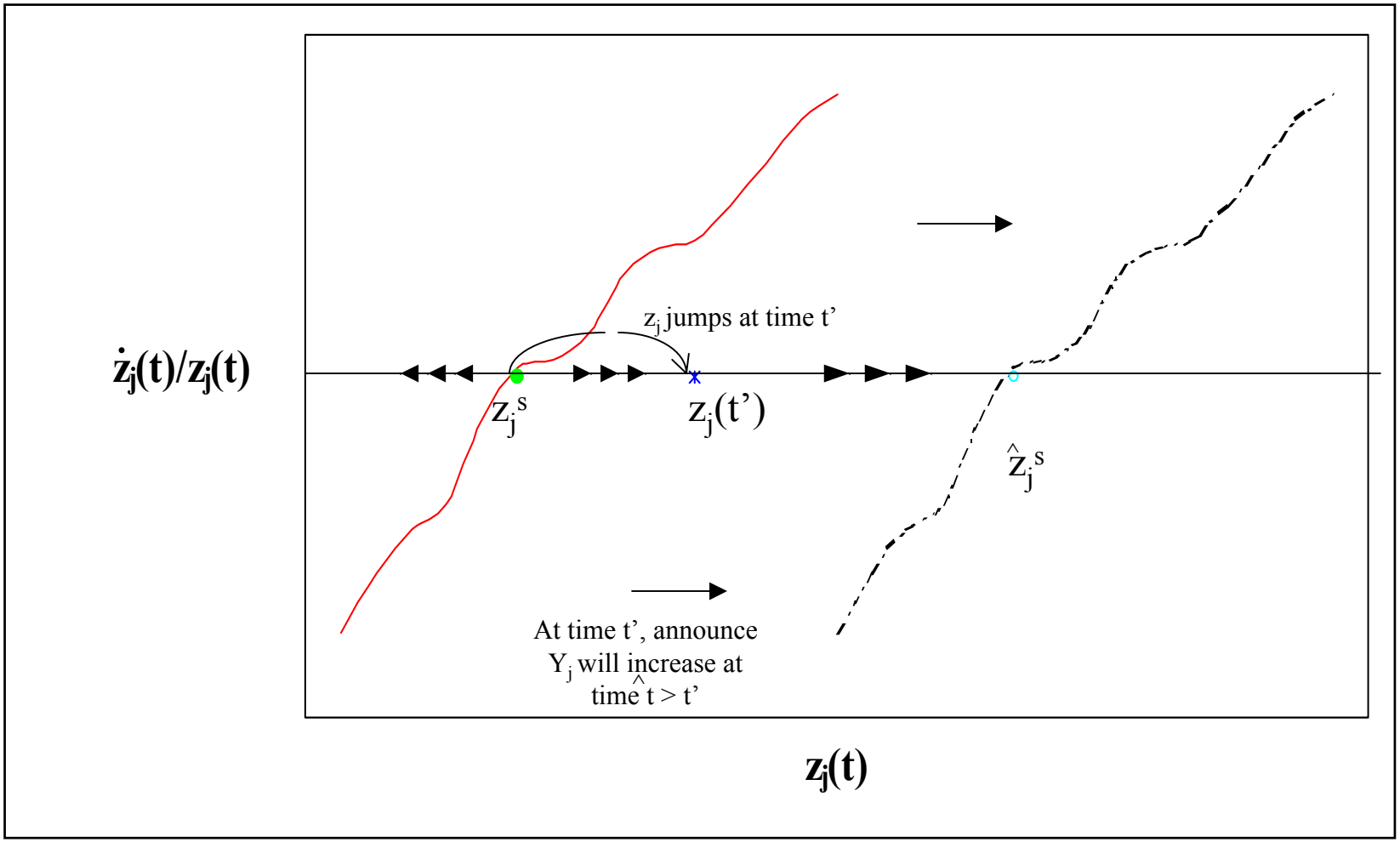


Figure 2: Share of Population by Age and Gender Groups, 1940-1990, from IPUMS

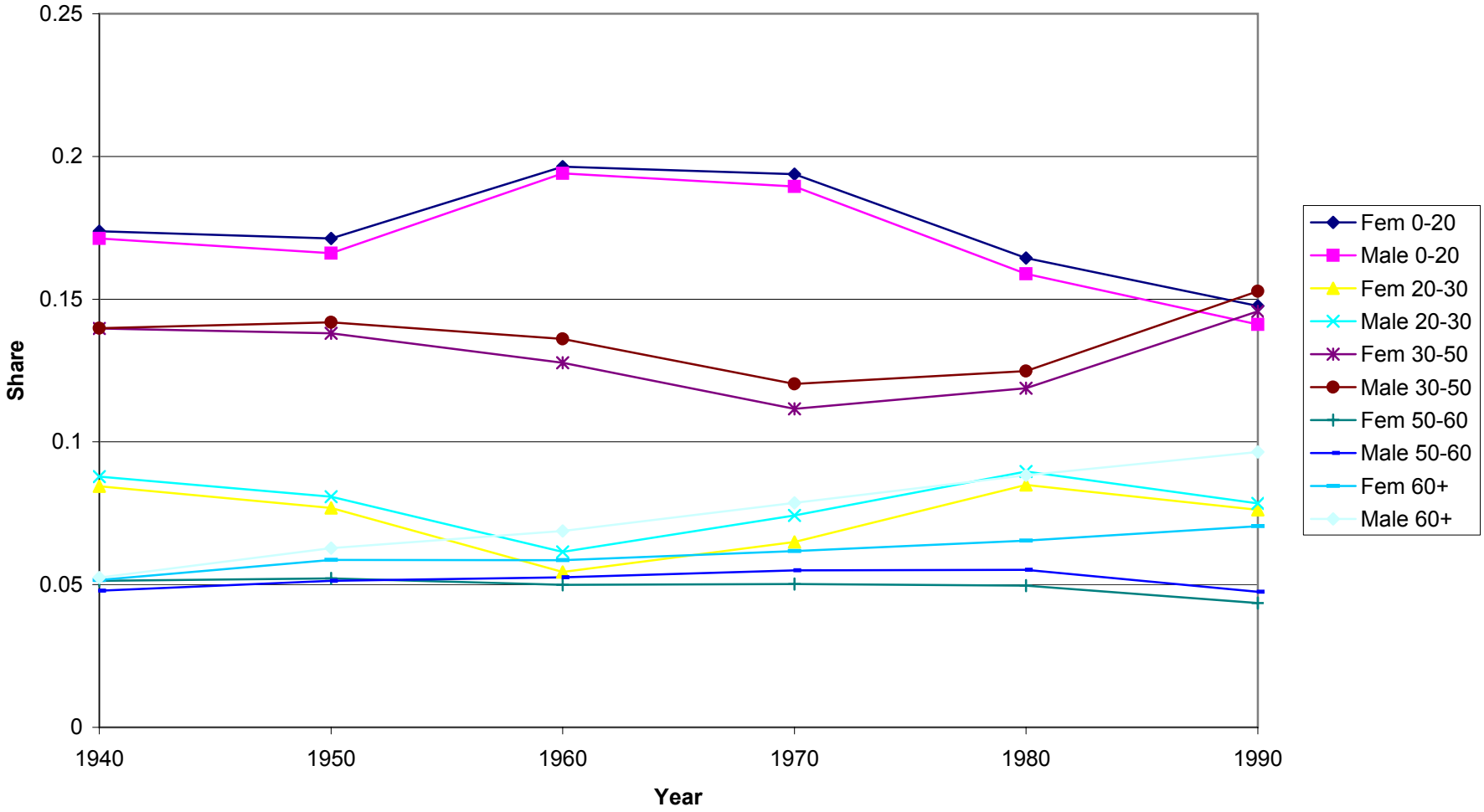


Figure 3: Share of Income by Age Group, 1970-2000, from CPS

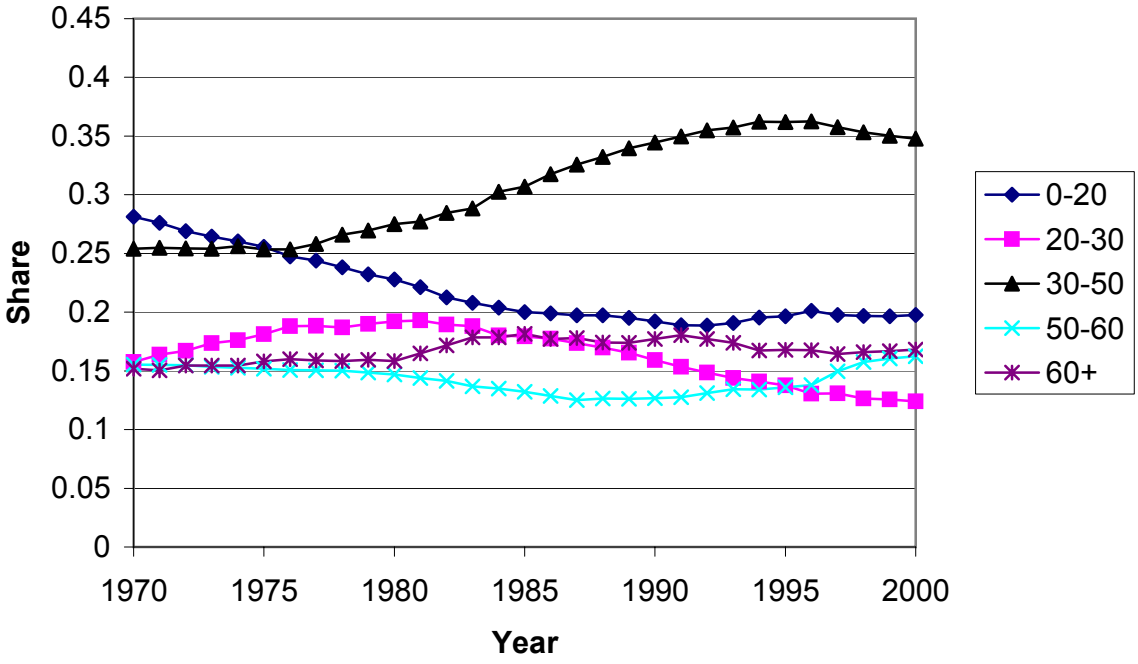
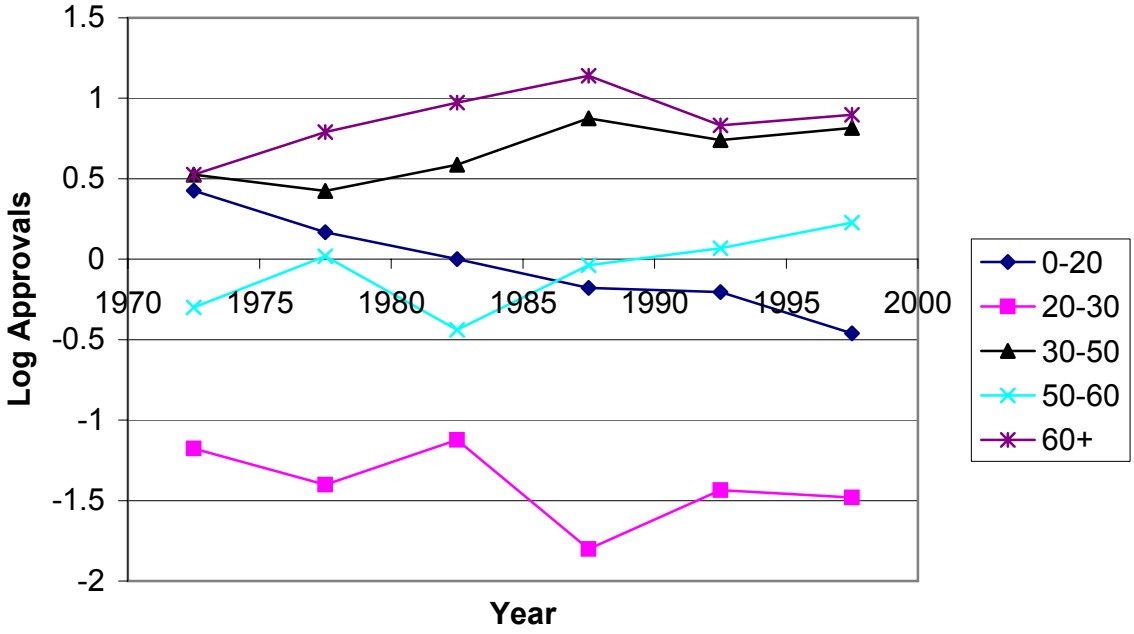
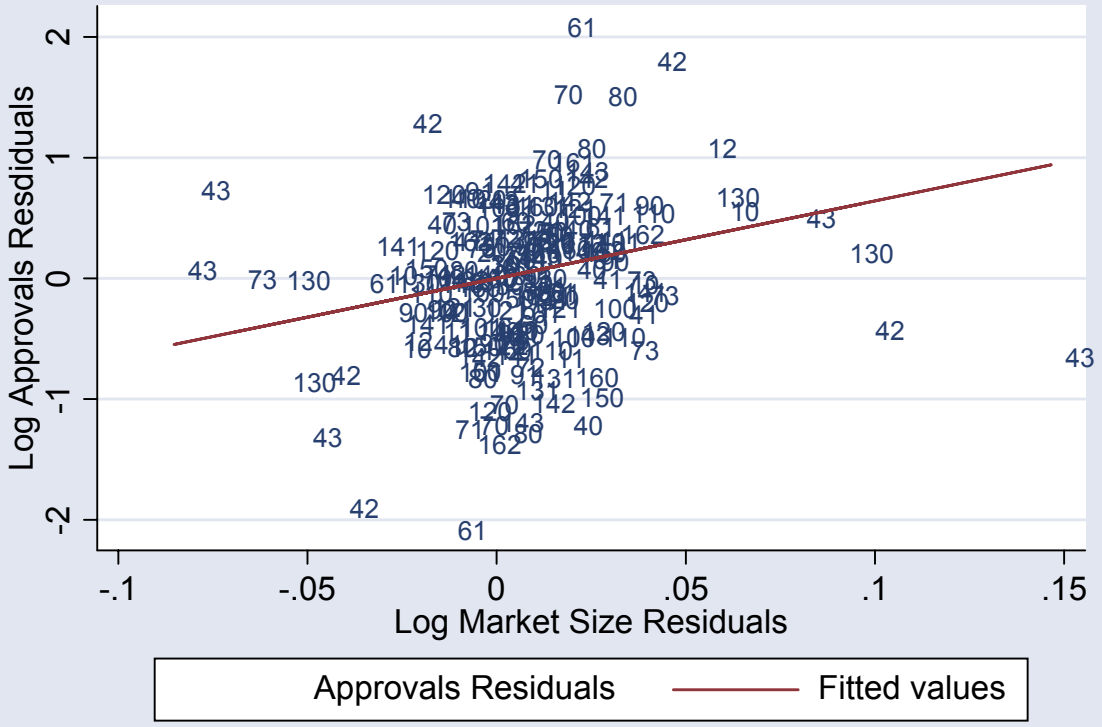


Figure 4: Log Drug Approvals by Age Group, from FDA



Notes: Drug approvals for each of the 34 categories are computed for five year invervals. The categories are combined into 5 groups, based on the age group that uses the category the most. Year dummies are removed from the approvals, and log approvals are plotted against time.

Figure 5: Approvals Residuals vs Market Size Residuals



Notes: Log approvals residuals and log market size residuals are residuals from OLS regressions of log approvals and log income-based market size on time and category dummies weighted by expenditure with five-year intervals. Fitted values are predicted log approvals residuals obtained from OLS regression in Table 3A column 1 panel B.