NBER WORKING PAPER SERIES

MARKET SIZE IN INNOVATION: THEORY AND EVIDENCE FROM THE PHARMACEUTICAL INDUSTRY

Daron Acemoglu
Joshua Linn

Working Paper 10038
http://www.nber.org/papers/w10038

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2003

We thank David Autor, Ernie Berndt, Olivier Blanchard, Alfonso Gambardella, Amy Finkelstein, Stan Finkelstein, Jerry Hausman, Paul Joskow, Frank Lichtenberg and participants at the NBER Summer Institute and Wharton 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. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Market Size in Innovation: Theory and Evidence From the Pharmaceutical Industry
Daron Acemoglu and Joshua Linn
NBER Working Paper No. 10038
October 2003
JEL No. O31, O33, L65

ABSTRACT

This paper investigates the effect of (potential) market size on entry of new drugs and pharmaceutical innovation. Focusing on exogenous changes driven by U.S. demographic trends, we find that a 1 percent increase in the potential market size for a drug category leads to a 4 to 6 percent increase in the number of new drugs in that category. This response comes from both the entry of generic drugs and new non-generic drugs, and is generally robust to controlling for a variety of non-profit factors, pre-existing trends, and changes in health care coverage.

Daron Acemoglu
Department of Economics
Massachusetts Institute of Technology
Cambridge, MA 02142
and NBER
daron@mit.edu

Joshua Linn
Department of Economics
Massachusetts Institute of Technology
Cambridge, MA 02142
jlinn@mit.edu

1 Introduction

This paper constructs a simple model linking innovation rates to current and future market size, and provides 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 important innovations 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 profit incentives and the size of the target market. For example, in his seminal study, *Invention and Economic Growth,* Schmookler argued that: "...invention is largely an economic activity which, like other economic activities, is pursued for gain" (1966, p. 206). To emphasize the role of market size, Schmookler entitled two of his chapters "*The amount of invention is governed by the extent of the market.*"

The role of profit incentives and market size in innovation is also important both 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), and for the induced innovation and directed technical change literatures, which investigate 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 and argue that there is generally insufficient research to develop cures for third-world diseases such as malaria.

In this paper, we investigate the effects of market size for different types of drugs on entry of new drugs and innovation. A major difficulty in any investigation of the impact of market size on innovation is the endogeneity of market size—better products will have bigger markets. Our strategy to overcome this problem is to exploit variations in market size driven by demographic changes (or past demographic trends), which should be exogenous to other, for example scientific, determinants of innovation and entry of new drugs.[2] We create the potential market

[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 that important developments took place despite the belief of many important figures in the development of the computer that there would not be a market greater than a handful of personal computers in the United States (2000, p. 13).

[2] For many drugs non-U.S. markets may also be relevant. Nevertheless, the U.S. market is disproportionately important, constituting about 40 percent of the world market (IMS, 2000). Below we also look at the impact of West European and Japanese demographic changes on the entry rates of new drugs.

size for various drug categories according to the age distribution of their users at a given point in time, and then trace changes in potential market size driven by changes in demographics (holding the age profile of consumption of various drug categories constant over time).[3] We measure entry and innovation using the Food and Drug Administration's (FDA) approval of new drugs.[4]

Our results show that there is an economically and statistically significant response of the introduction of new drugs to market size. For example, a 1 percent increase in the size of the potential market for a drug leads to a 4 to 6 percent increase in the number of new drugs. We check the robustness of our results by controlling for lagged FDA approvals, pre-existing trends, differences in the non-economic incentives to innovate, advances in biotechnology and changes in insurance coverage of drug expenditures. New drugs that enter the market comprise both generics and non-generics, and we find that both respond to market size, though the response of generics is larger and somewhat more precisely estimated.

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 size. On the other hand, because the development process of new drugs can be long, entry may respond to past market size. We find that current market size and 5-10 year leads of market size have the strongest effect on entry rates of new drugs, which is consistent with the predictions of the theoretical model we use to motivate our empirical investigation. Finally, we also look at the response of pharmaceutical patents to market size, and find a positive, but statistically insignificant relationship between future anticipated market size and patents, which might reflect the difficulty of matching patents to drug categories.

There are a number of other studies related to our work. First, Schmookler (1966) documents a statistical association between investments and sales, on the one hand, and patents and innovation, on the other, and argues 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. Pakes and Schankerman (1984) investigate this issue using a more structural approach, linking R&D intensity at the industry level to factor demands and to growth of output. In more recent research, Scott Morton (1999) and

[3] Loosely speaking, "market size" corresponds to the number of users times their marginal willingness to pay. Therefore, market size can increase both because the number of users increases and because of their marginal willingness to pay changes. We focus on changes driven by demographics to isolate exogenous changes in market size.

[4] These data were previously used by Lichtenberg and Virahbak (2002), who obtained them under the Freedom of Information Act. We thank Frank Lichtenberg for sharing these data with us.

Reiffen and Ward (2002) study the decision of firms to introduce a new generic drug and find a positive relationship between entry into a new market and expected revenues in the target market. None of these studies exploit a potentially exogenous source of variation in market size, however.

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, 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, which, they argue, was a response to higher energy prices. In a related study, Popp (2002) finds a strong correlation between aggregate patents and energy prices. These findings are consistent with the hypothesis that the type of innovation responds to profit incentives, though they do not establish causality.

Third, there is substantial research focusing on innovation in the pharmaceutical industry. 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. Ling, Berndt and Frank (2003) investigate the complementarity between new technologies and the skills of physicians in the spread of new drugs.

Most closely related to this study are Lichtenberg and Waldfogel (2003) and Finkelstein (2003). Lichtenberg and Waldfogel show that following the Orphan Drug Act there were larger declines in mortality among individuals with rare diseases (compared to other diseases) because of the incentives created by the Act to develop drugs for these rare diseases. Finkelstein 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 vaccinations, 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.[5]

The rest of the paper is organized as follows. We outline a simple model linking innovation to market size in the next section. Section 3 briefly explains our empirical strategy, and Section

[5] Lichtenberg (2003) also presents evidence suggesting that the types of new drugs changed towards drugs more useful for the elderly after Medicare was established.

4 describes the basic data sources and the construction of the key variables. Section 5 presents the empirical results and a variety of robustness checks. Section 6 contains some concluding remarks, while the Appendix gives further data details.

2 Theory

We now outline a simple framework to analyze the influence of market size on innovation. Subsection 2.1 discusses the basic model. Subsection 2.2 derives the implications of potential delays in the development and approval processes of new drugs. Subsection 2.3 generalizes the basic model to investigate the response of innovation effort and R&D to anticipated changes in future market size. Finally subsection 2.4 extends the model to introduce an explicit distinction between entry of generic drugs and non-generic drugs.

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 exogenously given endowment $y_i(t)$ at time t. Second, there is a large number J of drugs, x_1, ..., x_J. Each drug has a potentially time-varying "quality", $q_1(t)$, ..., $q_J(t)$. 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:

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

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.[6] 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

[6] Alternatively, suppose the preferences of individual i are:

$$\int_0^\infty \exp(-\omega(t, q_j(t) x_{ji}(t))) c_i(t)^{1-\gamma} dt,$$

where $\omega(t, q_j(t) x_{ji}(t))$ is the discount rate, which is influenced by the quality and quantity of drugs that the individual consumes (e.g., because they extend his life or affect its quality). If we assume that $\omega(t, q_j(t) x_{ji}(t)) = rt - \gamma \ln(q_j(t) x_{ji}(t))$, we obtain the expression in (1) in the text.

individual only consumes one type of drug are for simplicity and do not affect the main results.[7]

Normalizing the price of the basic good 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

$$x_{ij}(t) = \begin{cases} \frac{\gamma y_i(t)}{p_j(t)} & \text{for } i \in G_j \\ 0 & \text{for } i \notin G_j, \end{cases} \tag{2}$$

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. 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 (and is under patent protection indefinitely).

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. 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

$$n_j(t) = \delta_j z_j(t). \tag{3}$$

Differences in δ_j's introduce 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.

The most important feature of this R&D technology for our focus here is that technological progress is *directed* in the sense that firms can devote their research effort and expenditure to 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 a 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).[8]

[7] 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).

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.

[8] 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).

The demand curves in (2) have an elasticity equal to 1, so an unconstrained monopolist would like to charge an arbitrarily high price. However, the firm with the best drug in line j is 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 to exclude the next best firm—i.e., to ensure 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. Suppose that a consumer buys from the best firm with quality $q_j(t)$ and price $p_j(t)$ and chooses her optimal consumption as given by (2), then her instantaneous utility at time t will be:

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

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:

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

The limit price equalizes these two expressions, hence, equilibrium prices for all j and t satisfy:

$$p_j(t) = \lambda. \tag{4}$$

The profits of the firm with the best product of quality $q_j(t)$ in line j at time t are:

$$\begin{aligned} \pi_j(q_j(t)) &= (\lambda-1)\gamma\sum_{i\in G_j} y_i(t) \\ &= (\lambda-1)\gamma Y_j(t) \end{aligned} \tag{5}$$

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. Here $\lambda\gamma Y_j(t)$ corresponds to the market size (total sales) for drug j. Throughout we assume that all $Y_j(t)$'s are known in advance, which is plausible in the context of demographics-driven changes in demand. They can change because the number of consumers demanding this product changes, or because their incomes change, or also 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.

Firms are forward-looking, and discount future profits at the rate r. The discounted value of profits for firms can be written by a standard dynamic programming recursion. $V_j(t \mid q_j)$, the value of a firm that owns the most advanced drug of quality q_j in line j at time t, is:[9]

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \delta_j z_j(t \mid q_j) V_j(t \mid q_j) \tag{6}$$

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

for each $j = 1, 2, ..., J$, where $\pi_j(q_j)$ is the flow profits in drug line j given by (5), 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.

Note that entry of new drugs that are not technologically superior, but steal customers from the incumbent, such as the entry of generics, can also be included in $z_j(t)$, especially since equation (5) shows that the profits of a new entrant are 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 2.4, we present a model with separate entry rates of generics and non-generics.

There is free entry into R&D 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,

$$\text{if } z_j(t) > 0, \text{ then } \delta_j V_j(t \mid q_j) = 1. \tag{7}$$

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 (4), consumer demands for drugs $x_i(t)|_{i \in I}$ that satisfy (2) and sequences of R&D levels $z_j(t)|_{j=1,..J}$ that satisfy (7) with $V_j(\cdot)$ given by (6).

An equilibrium is straightforward to characterize. Differentiating equation (7) 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 (7) into (6) yields the levels of R&D effort in the unique equilibrium:

$$z_j(t) = \max\left\{\frac{\delta_j(\lambda - 1)\gamma Y_j(t) - r}{\delta_j}; 0\right\}, \tag{8}$$

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 $z_j(t) = (\delta_j(\lambda - 1)\gamma Y_j(t) - r)/\delta_j$, and we will often drop the max operator.

The most important feature of (8) 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 consequently, there will be greater research effort to acquire this position. Our empirical work below investigates 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 sufficient R&D to arbitrage them, thus ensuring $\dot{V}_j(t \mid q_j) = 0$. The intuition for the lack of 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.[10]

Combining equations (3) and (8) gives entry of new drugs as (ignoring the max operator):

$$n_j(t) = \delta_j(\lambda - 1)\gamma Y_j(t) - r. \tag{9}$$

This equation relates innovation or entry of new products, which we will approximate with FDA approval of new drugs, to market size (total expenditure of consumers in this line of drug). 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 market size may have only a small effect. Whether or not this is so is an empirical question.

[10] 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.

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 be as much as 15 years after 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. Therefore, we now have $n_j(t) = \delta_j z_j(t-T)$, where $z_j(t-T)$ is the rate of innovation at time $t-T$.

Given this structure, the key value function changes to:

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \delta_j z_j(t-T) V_j(t \mid q_j). \tag{10}$$

Equation (10) is a delayed differential equation rather than an ordinary differential equation, so a general analysis is more difficult. Nevertheless, the unique equilibrium in this case is still easy to characterize because of the simple structure here. Now the free entry condition is:

$$\text{if } z_j(t-T) > 0, \text{ then } \exp(-rT)\,\delta_j V_j(t \mid q_j) = 1, \tag{11}$$

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 (10) and (11) together imply that:

$$z_j(t-T) = \max\left\{\frac{\exp(-rT)\,\delta_j(\lambda-1)\gamma Y_j(t) - r}{\delta_j}; 0\right\}, \tag{12}$$

which is very similar to (8), 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 (12) may give the impression that there should now be a 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. Since $n_j(t) = \delta_j z_j(t-T)$, the key prediction of the model for entry of new drugs changes to:

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

which only differs from (9) by the term $\exp(-rT)$. This analysis therefore shows that delays in development and approval processes do not change the basic predictions relating to entry of new drugs, though the predictions about the timing of R&D and patenting are different (and this is useful to bear in mind when we look at patents below).

2.3 Anticipation Effects

We now generalize this basic setup to obtain a reaction to (anticipated) future market sizes. 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.[11] Given this specification, the value function changes to:

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \delta_j z_j(t)\phi(z_j(t))V_j(t \mid q_j), \tag{13}$$

for each $j = 1, 2, ..., J$, which only differs from (6) 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

$$\delta_j \phi(z_j(t))V_j(t \mid q_j) = 1, \tag{14}$$

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

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 (14) instead of (7) with $V_j(\cdot)$ given by (13).

To make further progress in this case, let us assume that $Y_j(t) = Y_j$ for all t, so that market sizes remain constant over time. Differentiating (14) with respect to time, and substituting into (13) and rearranging, we obtain J equilibrium differential equations in $z_j(t)$'s:

$$\frac{\dot{z}_j(t)}{z_j(t)} = \frac{1}{\varepsilon_\phi(z_j(t))}\left[r + \delta_j z_j(t)\phi(z_j(t)) - \delta_j \phi(z_j(t))(\lambda - 1)\gamma Y_j\right], \tag{15}$$

where $\varepsilon_\phi(z_j(t)) = -\phi'(z_j(t))z_j(t)/\phi(z_j(t))$ is the elasticity of the ϕ function.

[11] 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.

Since consumer incomes are now assumed to be constant, the steady-state equilibrium must have $\dot{z}_j(t) = 0$. This implies that the steady-state R&D level in drug line $j = 1, 2, ..., J$ is:

$$z_j^S = \frac{\delta_j \phi\left(z_j^S\right)(\lambda - 1)\gamma Y_j - r}{\delta_j \phi\left(z_j^S\right)}. \tag{16}$$

In the special case where $\phi(.) = 1$, this equation boils down to (8) in the basic model. 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 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 (15) 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, (15) 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 (16). 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, (14). This reasoning implies that there should be no anticipated jumps in $z_j(t)$, in particular no jump at $t = \hat{t}$, which 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 baseline model limit this response. In other words, even if a change in market size is anticipated far in advance, increasing R&D investment too far in advance would not be profitable because another firm is likely to innovate further before the actual increase in market size materializes. 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.[12]

[12] A previous version of the paper also investigated a number of other generalizations, which we omit because

2.4 Generics and Non-Generics

The analysis so far did not distinguish between generics and non-generics. In the empirical work, we first look at all entries, and then distinguish between generics and non-generic drugs—which 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, or they can engage in costly development to introduce a generic version of an already-existing drug (without quality improvement).[13] 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).[14]

Let us suppose that one unit of the final good devoted to prepare a generic for the market in drug line j leads to successful entry at time t at the flow rate θ_j, so entry of new generics at time t is given by $g_j(t) = \theta_j h_j(t)$, where $h_j(t)$ is total generic development expenditure for drug line j at time t. In practice, in addition to technological factors, length of patents on non-generics influences θ_j. We assume that $\theta_j > \delta_j$, since introducing a generic to the market must be easier than inventing a new drug.

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 (5), 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 (e.g., Cournot) competition or collusion, so $\mu > 0$ is possible. If

of space constraints.

Briefly, we allowed research to be only imperfectly directed, so that research towards 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 $(1-p)\delta_{j'}/J$ any $j' = 1, ..., J$, where $1 \geq p > 0$. When $p = 1$, we have the model of subsection 2.1, while with $p \to 0$, research becomes undirected. Interestingly, this generalization does not affect the results of the model, and even with $p \to 0$, the innovation rates are given exactly by (9), and thus the results do not depend on the assumption of perfectly directed research.

We also generalized to set up by allowing "technological drift", meaning random innovations arising from non-profit and non-economic motives, which again did not affect the results, and led to an equilibrium relationship similar to (9). Details are available upon request.

[13] An alternative approach would be to link the entry of non-generics directly to the delayed entry of generics, since generics can only enter after the patents on previous non-generics expire. Nevertheless, because introducing a generic into the market still involves significant upfront costs, we expect profit incentives to play a major role here. A hybrid model incorporating both such delays and profit incentives is significantly more complicated to analyze.

[14] Prior to 1984, the FDA demanded a similar application process for approval of a generic drug as for a non-generic. For example, the generic drug's safety and efficacy had to be demonstrated through clinical trials. DiMasi et al (1991) suggest that the cost of approval for a generic was as much as 40% of the cost of a non-generic. The Hatch-Waxman Act in 1984 allowed a company to obtain FDA approval for a generic drug by demonstrating that it is molecularly similar to and has the same active ingredients, indications and strength as the non-generic, substantially reducing the costs of introducing a generic. Reiffen and Ward (2002) estimate the cost of introducing a new generic to be around \$5 million after the Act, compared to over one hundred million dollars before (DiMasi et al, 1991).

$\mu = 0$, 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 for further entry of this sort introduces additional notation, but does not affect our qualitative results.

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

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j) - \delta_j z_j(q_j) V_j(t \mid q_j) - \theta_j h_j(t) [V_j(t \mid q_j) - W_j(t \mid q_j)], \quad (17)$$

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 (6) above is the last term, which takes into account that at the flow rate $\theta_j h_j(t)$ 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, this value is given by:

$$rW_j(t \mid q_j) - \dot{W}_j(t \mid q_j) = \mu\pi_j(q_j) - \delta_j z_j(t) W_j(t \mid q_j). \quad (18)$$

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

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

These conditions are similar to (7) above, but written out 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 next assume that

$$\mu\theta_j < \delta_j \quad (19)$$

for all j. If this assumption does not hold, entry of new generics is so rapid that it ceases to become profitable for pharmaceutical companies to introduce new drugs, and consequently, quality improvements come to an end. As long as Assumption (19) holds, similar arguments to before imply that the unique equilibrium is given by:

$$\begin{aligned} z_j(t) &= \frac{\mu\theta_j(\lambda - 1)\gamma Y_j(t) - r}{\delta_j}, \quad (20) \\ h_j(t) &= \frac{(\delta_j - \mu\theta_j)(\lambda - 1)\gamma Y_j(t)}{\theta_j - \delta_j}. \end{aligned}$$

It can also be verified that if $\mu = 0$, the equilibrium of Proposition 1 applies, and if Assumption (19) does not hold, there will be no R&D, i.e., $z_j(t) = 0$, and thus $\lim_{t\to\infty} h_j(t) = 0$ (or $\Pr \lim_{t\to\infty} h_j(t) = 0$) as there will eventually be two producers in the market for drug j.

Moreover, the entry rate of non-generics and generics, $n_j(t)$ and $g_j(t)$, are:

$$\begin{aligned} n_j(t) &= \mu\theta_j(\lambda - 1)\gamma Y_j(t) - r, \\ g_j(t) &= \frac{\theta_j(\delta_j - \mu\theta_j)(\lambda - 1)\gamma Y_j(t)}{\theta_j - \delta_j}. \end{aligned} \tag{21}$$

There are a number of important points to note about this equilibrium. First, 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 (21) 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.[15]

3 Empirical Strategy

3.1 Empirical Specification and Estimation Issues

As $r \to 0$, we can take logs on both sides of equation (9) to obtain:

$$\log n_j(t) = \text{constant} + \log \delta_j + \log m_j(t), \tag{22}$$

where $m_j(t) \equiv \lambda\gamma Y_j(t)$ is the market size for drug line j at time t. We measure entry of new drugs (or innovation), $n_j(t)$, as new drug approvals by the FDA (Food and Drug Administration) in broad drug categories as described below.[16] This measure, denoted by N_{ct} for drug category c at time t, includes entry of generic drugs. Although generic drugs do not correspond to "innovation" according to the standard use of this term, they are still driven by the same

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

[16] Besides new drug approvals and patents, which we also use below, there is a third proxy for innovation rates: clinical trials. We were unable to obtain data on clinical trials for a sufficient number of drug categories.

profit incentives as innovation, and are similar to innovation in the context of our model.[17] 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.

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

$$\log N_{ct} = \alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \zeta_c + \mu_t + \varepsilon_{ct}, \tag{23}$$

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, ζ_c's are a full set of category fixed effects that correspond to the $\log \delta_j$ terms above, μ_t's are a full set of time effects capturing any common time component, and finally, ε_{ct} is a random disturbance term, capturing all omitted influences. The specification with the dependent variable in logarithm is useful, since it ensures that drug category fixed effects and time effects have proportional impacts on entry of new drugs.

One problem with equation (23) 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 it impossible to estimate (23). We take a number of approaches to this problem. First, we change (23) to

$$\log \tilde{N}_{ct} = \alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \gamma \cdot d_{ct} + \zeta_c + \mu_t + \varepsilon_{ct}, \tag{24}$$

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, which was first used by Pakes and Griliches (1980), has the advantage of simplicity and flexibility (the data determine how $N_{ct} = 0$ should be treated). The drawkback is that the variable d_{ct} is mechanically a function of N_{ct}, so it can introduce various biases.

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

$$N_{ct} = \exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \zeta_c + \mu_t) + \varepsilon_{ct}, \tag{25}$$

[17] Moreover, if generics and non-generics are imperfect substitutes, generics would correspond to new products, and their entry would increase consumer welfare in a manner similar to technological improvements, and in fact, formally correspond to technological change in the product variety models (e.g., Romer, 1990, Grossman and Helpman, 1991).

which is a slight variant on equation (23) above. Estimates from the two models are typically similar 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 are larger discrepancies between the two models, and in those cases we favor the Poisson model.

The estimation of (25) would lead to biased estimates, however, since the fixed effects, the ζ_c's, cannot be estimated consistently. To deal with this problem, we follow Hausman, Hall, and Griliches (1984), and transform (25) to obtain:

$$S_{ct} = \frac{\exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \mu_t)}{\sum_{\tau=1}^{T} \exp(\alpha \cdot \log M_{c\tau} + X'_{c\tau} \cdot \beta + \mu_\tau)} + \varepsilon_{ct}, \tag{26}$$

where $S_{ct} = N_{ct} / \sum_{\tau=1}^{T} N_{c\tau}$ 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. This transformation removes the drug category dummies, and the coefficient of interest, α, can be estimated consistently. We estimate this equation using nonlinear least squares (NLLS) as well as maximum likelihood (ML). Woodridge (1999) shows that NLLS estimation strategy has good consistency properties, even when the true model is not Poisson. Finally, we also report some estimates from the negative binomial model which relaxes the distributional assumptions of the Poisson model in a maximum likelihood context (see, for example, Wooldridge, 2002, or Hausman, Hall, and Griliches, 1984).

In addition to equations (24) and (26), which have the contemporaneous value of $\log M_{ct}$ on the right hand side, we also estimate equations with leads and lags of $\log M_{ct}$ to determine whether there are significant delays and anticipation effects. Such effects are possible, since, as reported by DiMasi et al. (1991), it can take as long as 15 years for a drug to enter the market from the time of initial research. Furthermore, changes in 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.

3.2 Potential Market Size and Identification

Throughout, we exploit the potentially exogenous component of market size driven by demographic trends, combined with differences in the age distribution of users for different types of drugs. We obtain the age composition of users (and expenditure) from micro drug consumption data, and the changes in U.S. demographics from the CPS (Current Population Survey) data. Our measure of potential market size is:

$$M_{ct} = \sum_{a} u_{ca} \cdot p_{at}, \tag{27}$$

where p_{at} is the U.S. population or income of age group a at time t, and u_{ca} is the time-invariant measure of consumption of drug category c by individuals in age group a. We take time periods to be either five-year or ten-year intervals.

We compute M_{ct} in two alternative ways: first, we use the U.S. population for age group a at time t for p_{at} and the average number of drugs in category c used per person in age group a for u_{ca}; and second, we use the total income of age group a for p_{at} and the average share of drugs in category c in the total expenditure of those in age group a for u_{ca}. The latter, which we refer to be as 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.[18] For both measures, the over-time source of variation is not from changes in individual use, but purely from demographic changes captured by p_{at}—i.e., u_{ca}'s are not time-varying. So for example, changes in prices, which potentially result from innovations and affect consumption patterns, will not cause over-time variation in M_{ct}.

The major threat to the validity of our empirical strategy is from potentially time-varying omitted variables (any variable that is not time-varying is taken out by the drug category fixed effects). Omitted variables related to market size or profit opportunities may induce a bias in the implied magnitudes, but will not lead to spurious positive estimates of the effect of market size (in other words, the presence of such variables is essentially equivalent to mismeasurement of the appropriate market size). 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 entry. We attempt to substantiate our identifying assumption further by including lagged dependent variables, and adding controls such as pre-existing trends and proxies for other incentives to undertake research in a particular field.[19]

4 Data and Descriptive Statistics

4.1 Basic Data Sources

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 drug use patterns of these age groups. To construct income shares, we divide household income equally among the

[18] If preferences have a Cobb-Douglas form as in (1) and are stable, the expenditure measure of u_{ca} will always be constant. With non-Cobb-Douglas preferences, changes in prices will induce changes in u_{ca} over time. In this case, by using a time-invariant measure of u_{ca}, we remove this potentially endogenous source of variation.

[19] Another source of endogeneity may be that innovations in certain drug categories extend the lives of the elderly, thus increasing their M_{ct}. Lichtenberg (2003) provides evidence that new drugs extend lives. This source of endogeneity is not likely to be quantitatively important, however, since the variation resulting 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} with past demographics, purging it from changes in longevity.

members of the household. Figure 2 shows population shares for the five age groups, and Figure 3 shows the corresponding income shares (i.e., income of the corresponding age group divided by total income). To facilitate comparison with Figure 4, this figure starts in 1970. Both 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.

The FDA classifies all prescription drugs into 20 major drug categories, which are further subdivided into 159 categories. These categories are based on a combination of therapeutic intent and chemical structure.[20] We drop 4 of the 20 major categories from this classification: Anesthetics, Antidotes, Radiopharmaceuticals and Miscellaneous.[21] We obtain a total of 34 categories by breaking 10 of the 16 remaining categories into finer groups when there are significant heterogeneity in terms of the age distribution of users across subcategories. We separate the major categories of Antimicrobials, Psychopharmacologics, Nutrients, Hormones, Dermatologics, Neorologics, Ophthalmics, Otologics, Pain Relief and Respiratory because within these categories there are 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 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 per person and expenditure share for each category and each of our five age groups. Appendix Table A1 reports these numbers, and 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, and Penicillins and Antibacterials are used most by individuals in the youngest group. We construct the measures of potential market size according to equation (27) by combining data from the MEPS and the CPS.

We supplement the MEPS data with the National Ambulatory Medical Care Survey (NAMCS),

[20] Other authors, for example Lichtenberg (2003), use a different classification system based on diseases. For our purposes, the FDA classification is attractive, both because it enables us to construct a better classification of all approved drugs currently on the market, and because the microdata surveys give the FDA classes for the drugs that the respondents use (see Data Appendix for details).

[21] We drop the Anesthetics, Radiopharmaceuticals and 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 mainly contains surgical and dental tools. The Antidote category is dropped because there are few drugs approved and there is little use of these drugs in the surveys. See the Data Appendix for further details on the contruction of our categories.

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 week. The main use of the NAMCS 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.[22]

Table 1 gives correlations between various measures of drug use. Panel A shows a high degree of correlation between the NAMCS surveys at various dates, indicating that the 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 MEPS or NAMCS. Weighted correlations are larger than the overall correlation because our estimates of the age distribution of users are less precise when there are fewer individuals using drugs in a particular category. 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 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 substantial persistence in the age distribution of drug expenditure. Finally, panel C shows a high degree of correlation between expenditure shares 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 new FDA drug approvals. We exclude over-the-counter drugs, the so-called orphan drugs,[23] and drugs that have the same identifying characteristics

[22] Appendix Table A2, which is available upon request, compares this classification system with the 34 category system developed with the MEPS. Some of the 159 categories have been dropped from one classification system, but not from the other, because there were not sufficient observations to construct reliable estimates of drug use from one of the surveys. Appendix Table A2 shows that the two systems are closely related. In general the 34 category system is a slightly less aggregated version of the 30 category system. Nevertheless, 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).

[23] These drugs 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 because they receive special inducements under the Orphan Drug

(i.e., same name, company, and category, or the same FDA approval number). We focus on the time period 1970-2000. Since we can only match FDA categories for drugs that are still listed by the FDA, the quality of the approvals data deteriorates as we go back in time. In addition, because we are using age composition from the 1990s, the quality of our measures of potential market size also deteriorates as we go back in time. Our approvals dataset for 1970-2000 comprises 6,595 prescription drugs, including both generics and non-generics (see the Appendix). Since 1970 there have been about half as many approved non-generics as generics.

Figure 4 shows the share 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 accounts for the largest fraction of use in that category (thus this cut of the data uses only part of information that we will exploit in the regression analysis), and compute the share of drug approvals by dividing the number of approvals in a given category by total approvals in that time period.[24] 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 income share of the 30-50 group increases over the sample, and so does the entry of drugs most used by this group. Both the shares of income and entry of drugs for those 0-20, on the other hand, show a downward trend, while those for the 60+ age group show an increase followed by a decline. Table 2 also shows changes in income shares by age group and in FDA drug approvals (for all drugs as well as separately for generics and non-generics), and confirms the patterns depicted in Figures 3 and 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 (24) and (26) 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

Act.

[24] There are large fluctuations in the total number of approvals, presumably because of a number of institutional changes. For example, it was discovered in 1989 that some FDA officials were taking bribes to speed up the approval process for generic drugs. As a result, in the early 1990s 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 1990s, but only a small decline for non-generics. We thank Ernie Berndt for suggestions on this issue.

other than drug category effects and time effects. The results show a large and significant effect of (potential) market size ($\log M_{ct}$) on entry of new drugs.

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 the time periods corresponding to five-year intervals. Since our estimates of age composition in smaller categories are less precise, observations are weighted by MEPS cell sizes (total expenditure for income-based measures in the cell and total drug use for population-based measures). The standard errors throughout are corrected for heteroscedasticity using the Huber-White formula. The NLLS estimate of α in this basic specification is 6.04 with a standard error of 1.95, which is significant at the 1 percent level. The OLS estimate is somewhat smaller, 4.52, with standard error 2.00, thus significant only at the 5 percent level.[25]

Potential market size appears to explain a sizable fraction of the total variation in the entry of new drugs (for example, the partial R^2 of the OLS specification is 0.20), and the empirical relationship is not driven by outliers. 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. Although 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, excluding these has little effect on our estimate of α. For example, dropping these three categories leads to an NLLS estimate of 7.07 (s.e.= 2.29).

The quantitative magnitude of the effect in column 1 is large but plausible, implying that a 1 per cent increase in our market size measure leads to about a 6 percent increase in drug approvals. Since there are a total of 6,595 approvals between 1970 and 2000, thus on average 32 approvals in every five-year interval in each of our 34 categories, the estimate of 6.04 implies that a 1 percent increase in market size leads to the entry of about two new drugs. Total pharmaceutical sales was approximately \$130 billion in 1999 (IMS, 2000), which implies an average annual expenditure of \$3.8 billion per category. A 1 percent increase therefore corresponds to \$38 million, or about \$570 million over 15 years, which is the life of a typical drug. Since entry costs for non-generics are around \$800 million (in 2000 dollars, DiMasi et al, 2001), while for generics they have varied between \$5 million and over \$100 million (DiMasi et al, 1991, Reiffen and Ward, 2002), entry of two new drugs in response to an increase of \$570 million in revenue

[25] Clustering the standard errors at the level of drug categories has little effect because there is no residual serial correlation—see the lagged dependent variable specifications reported in Table 6.

is within the range of plausible responses.

In column 2, we use ten-year intervals instead of the five-year intervals, both to reduce the potential noise in the entry of new drugs during the five-year intervals and also to check whether ten-year intervals correspond more closely to the relevant match between market size and entry. The estimate of α using NLLS is now 4.95 with standard error 1.90, significant at the 1 percent level. The smaller NLLS estimate indicates that there might be some attenuation with the ten-year invtervals, but the ten-year OLS estimate is slightly larger than the five-year estimate. In the remainder, we mainly focus on five-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 slightly larger for both NLLS and OLS, and are estimated with about the same level of precision.[26]

To see the effect of weighting on the estimates, columns 5 and 6 report unweighted regressions. Without weights, the NLLS estimates are smaller than in columns 1 and 2, but still significant at the 1 percent level, and the OLS estimates, also smaller than in columns 1 and 2, are significant at the 5 percent level. We conjecture that this somewhat weaker relationship without weights reflects the less precise estimation of age composition in the smaller categories, which is illustrated by the correlations in Table 1.[27]

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 to estimate equation (26). 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 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 similar estimates to those in columns 1-4 of Table 3A. Finally, we also report estimates from the negative binomial model in panel C, which allows for overdispersion of the Poisson parameter. In this case, the estimates of α are slightly smaller than those in columns 1-4 of Table 3A, and

[26] We also estimated these models using West European and Japanese demographic information in addition to U.S. information (obtained from the United Nations website, esa.un.org/unpp/). Using the total population-based market size measure combining European, Japanese and U.S. populations gives similar results to those obtained using only U.S. information. For example, the NLLS specification using this total market size and five-year intervals gives an estimate of 5.27 (s.e.=1.50), as compared to the corresponding estimate of 6.16 in column 3.

[27] This conjecture also receives support from the fact that when we construct our potential market size measure using the NAMCS, which has a more even distribution of observations across drug categories, the unweighted and weighted results are similar (see Table 7 below).

still significant at 1 or 5 percent.

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 anticipation effects, especially since demographics-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 equations.

Columns 1 and 2 of Table 4 replicate the basic five-year and ten-year specifications from Table 3A for comparison. Columns 3 and 4 include the market size from the previous period and show that new drug entry responds mainly to current market size (in fact, the coefficient on lag market size is negative, though insignificant, presumably because current and previous market sizes are highly correlated). For example, in column 3, the coefficient on current market size is larger than our baseline, 13.56 (s.e.=4.45) with NLLS, and 8.62 (s.e.= 3.91) with OLS.[28] Columns 5 and 6 show that lag market size is also typically insignificant when entered by itself (with the exception of Panel A in column 5, where it is significant but much smaller than the estimate with current market size in column 1). These results therefore show no evidence of significant delays in the response of new drug entries to changes in market size.

In column 7, we include the current and one period ahead market size. Both the contemporaneous and lead market sizes are individually insignificant, but jointly significant at the 1 percent level. The coefficient on lead market size is much larger. In column 8, when we use ten-year intervals, instead, the coefficients are about the same magnitude, but current market size is significant at the 1 percent level (and similar to the baseline) while the lead market size is insignificant. Columns 9 and 10 show large and significant coefficients when we include only future market size.[29] These results therefore suggest that there may be some anticipation effects, perhaps with five- or ten-year leads, which is consistent with the possibility of limited anticipation effects highlighted by the theoretical model.[30]

[28] We construct the lagged market size measures for 1960s using demographic information from the CPS, so the number of observations does not decline. The results are similar if we only use the post-1970 data.

[29] We have also extended the sample for the lead specification by combining the population projections of the U.S. Census Bureau for 2010 with the incomes from the 1990s (the results are also similar if we use a linear extrapolation to predict future income). Using this additional period yields similar results. For example, re-estimating the specification in column 4 with NLLS, the estimate on current market size is 3.10 (s.e.=3.75) and the estimate on lead market size is 4.33 (s.e.=4.03), jointly significant at 1 percent. The specification in column 6 gives an estimate on lead market size of 7.10 (s.e.=1.86).

[30] Here "limited" does not refer to the strength of the effect, but to the fact that the response to market size

5.3 Potential Supply-Side Determinants of Innovation

In this subsection, we investigate the robustness of the baseline results to controlling for potential non-profit determinants of innovation, such as changes in scientific incentives or opportunities captured by the δ_j's in the theoretical model.

First, recall that the major threat to our identification strategy is changes in the δ_j's (since permanent differences in δ_j's are already taken out by our drug category fixed effects). If the δ_j's change over time, they are also likely to be serially correlated. Adding lags of $\log N_{ct}$ to our basic specifications is therefore a simple way to check for the importance of these concerns.

Columns 1 and 2 of Table 5 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:

$$\log N_{ct} = \alpha \cdot \log M_{ct} + \psi \cdot \log N_{ct-1} + \gamma \cdot d_{ct} + \delta_c + \mu_t + \varepsilon_{ct}. \tag{28}$$

Since $\log N_{ct-1}$ is correlated with the error term mechanically, 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 additional autocorrelation in the error term, ε_{ct} (see, for example, Blundell and Bond, 1998). This specification is also useful more generally to check for other sources of serial correlation in the entry rate of new drugs (such serial correlation would make the arguments for the exogeneity of $\log M_{ct}$ less compelling).

The estimates of α from equation (28), reported in columns 1 and 2, are quite similar to the baseline. The coefficient on the lagged dependent variable, $\log N_{ct-1}$, is essentially 0 and insignificant for five-year intervals; for ten-year intervals, though significant with OLS, it is again insignificant with NLLS. These results therefore show that there is no significant residual serial correlation, due to either changes in scientific opportunities or other sources, and that controlling for such serial correlation has no effect on our estimates.

A plausible conjecture is that non-profit incentives to develop drugs would be particularly responsive to opportunities to save lives or cure major illnesses. Motivated by this reasoning, 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 (see Lichtenberg, 2003, on the effect of pharmaceutical innovations on declines in mortality). To investigate this issue, we measure the number of

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 1.57 (s.e.=2.52), and the OLS estimate is 1.16 (s.e.=6.06).

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 total number of life-years lost for all the deaths resulting from diseases related to drugs in each category.[31]

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. 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 5, and leads to an estimate of 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.[32]

Starting in column 3, in addition to NLLS estimates in panel A and OLS estimates in panel B, which use current market size, we also report NLLS estimates with leads of market size in panel C. These specifications typically yield results very similar to those in Table 4.

Third, we investigate the implications of differences in scientific funding for various drug categories. Using the Computer Retrieval of Information on Scientific Projects (CRISP) dataset (details in the Data Appendix), we construct a variable measuring the total amount of federal funding for research projects in all drug categories, and 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 plan 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 estimates of α.

Fourth, to control for potential trends in scientific opportunities across drug categories, we add proxies for pre-existing trends. We construct an estimate for pre-existing trends as $\Delta_c = (\log N_{c,60} - \log N_{c,40})/2$, where $\log N_{c,60}$ is the log approvals for category c in 1960 and $\log N_{c,40}$ is the log approvals in 1940. We then estimate the equation:

$$\log N_{ct} = \alpha \cdot \log M_{ct} + \sum_{i=80,90} \Delta_c \cdot \sigma_i + \delta_c + \mu_t + \varepsilon_{ct}, \tag{29}$$

where σ_i's are dummies for the 1980s and 1990s. This specification allows drug categories that

[31] For example, if someone dies at age 32, this counts as 33 life years lost; people dying older than 65 receive no weight in this calculation.

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 a number of empty cells.

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

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 estimate of α is similar to our baseline estimate, 6.23, with standard error 1.88. 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. The resulting estimate is again similar to our baseline. These results are perhaps not surprising, since pre-1970 approvals are considerably noisier, thus only an imperfect control for pre-existing trends.

An alternative, and substantially more demanding, strategy is to include linear time trends. To do so, we estimate:

$$\log N_{ct} = \alpha \cdot \log M_{ct} + \eta_C \cdot t + \delta_c + \mu_t + \varepsilon_{ct}, \tag{30}$$

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 major drug categories each drug belongs to, since these categories are based on therapeutic intent, while the subcategories are based on use by age group. The estimates, reported in column 9, are close to our baseline. For example, with NLLS, the estimate of α is 5.60 (s.e.=2.10). The OLS estimate is smaller and insignificant. Using lead market size instead of current market size results in a similar pattern: the NLLS estimate, 8.46 (s.e.=2.53), is significant at 1 percent, while the OLS estimate (not reported to save space), 4.99 (s.e.=2.81), is significant at 10 percent.

We also investigate 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 witnessed the entry of the greatest number of orphan drugs, presumably by biotechnology firms. Although, as noted above, our dependent variable does not include these drugs, we also check whether our results are driven by entry of new drugs in these categories. The estimates in column 10 are close to those in column 1 of Table 3A, demonstrating that the estimates of α 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.[33] In column 11, we drop these two categories, and again find that our results are essentially unchanged.

To assess the role of biotechnology firms further, we add the approvals of a group of products known as biologics, where biotechnology firms have been active, to our measure of drug

[33] Biotechnology firms were also active in producing human growth factor, but since there are only a small number of individuals using these drugs in the MEPS and NAMCS, these drugs are not included in our approvals dataset.

approvals. These products include some vaccines, blood and plasma related products, and other products such as interferon and erythroproteins (used for red blood cell production), and are not included in our baseline measure because they go through a separate FDA regulatory process. The results of this regression, reported in column 12, show little change in the estimates of α.[34]

Finally, to see whether the advent of biotechnology or other technological advances of the past two decades have changed the relationship between market size and entry of new drugs, we estimated our baseline models including an interaction between a post-1985 (or post-1990) dummy and market size. Our estimates showed no evidence of significant interactions. For example, in a specification parallel to the NLLS model of column 1 of Table 3A, the estimate of α is 5.24 (s.e.=1.96), and the interaction with the post-1985 dummy is 0.10 (s.e.=0.07), thus quantitatively very small and insignificant.

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 findings. Although these results are not conclusive on the effect of scientific or other non-profit considerations in pharmaceutical research, they suggest that the effect of potential market size on entry and innovation is relatively robust.

5.4 Changes in Health Insurance Coverage

Our market size measure only exploits changes in potential market size driven by demographic trends. Another 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 68% between 1974 and 1996, while the percentage of 60+ year-olds with private insurance rose from 62% to 75% (authors' calculations from the National Health Interview Survey). Furthermore, there have been changes in Medicaid eligibility rules, designed to insure more poor children. These changes in health insurance coverage induce additional changes in market sizes. We now investigate both whether controlling for this source of variation affects the estimates of the impact of the demographics-driven potential market size measure, M_{ct}, and whether we can improve our measure of potential market size by including information on health insurance coverage.

[34] We also repeated our basic regressions without the categories where biologics are most common, Antivirals (12), Hematologics (20) and Immunologics (80), with little effect on the estimates of α.

We use the National Health Interview Survey (NHIS, 1974-1996) 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 and Medicare. Because Medicare does not cover prescription drugs, it enables us to perform a simple "falsification test".

First, with direct parallel to our M_{ct} measure, we use the NHIS to construct the following insurance coverage variable:

$$\bar{H}_{ct} = \sum_a u_{ca} \cdot \bar{p}_a \cdot f_{at}, \tag{31}$$

where f_{at} is the fraction of age group a in period t with private health insurance, u_{ca} is the expenditure share as described above, and $\bar{p}_a$ is the sample average income of age group a (in this section, we always use income-based market size and weights). We use $\bar{p}_a$ rather than p_{at} to isolate changes due to drug insurance coverage rather than demographic changes. We then add $\log \bar{H}_{ct}$ to our estimating equations (24) and (26).

Column 1 of Table 6 shows that the addition of this variable has no effect on the estimates of α relative to those in Table 3A, and that our baseline results are robust to controlling for separate trends in private health-care coverage. For example, in the NLLS specification the coefficient is now 5.21, as opposed to 6.04 in Table 3A. The coefficient on $\log \bar{H}_{ct}$ itself is positive, though insignificant.

Next, we construct alternative measures of $\log \bar{H}_{ct}$ using Medicaid and Medicare coverage rates for f_{at} (though all individuals above 65 are covered by Medicare, not all of them take up the benefits). Not surprisingly, with Medicaid and Medicare, the estimates of α are unaffected and the effect of $\log \bar{H}_{ct}$ is substantially weaker, presumably because these measures exploit much less of the variation in actual market size. Columns 4-6 investigate the implications of adding the $\bar{H}_{ct}$ measures in the specifications with lead market size, and show similar results.

Columns 7-9 take an alternative approach and construct a market size measure incorporating both demographic changes and changes in insurance coverage:

$$\widetilde{H}_{ct} = \sum_a u_{ca} \cdot p_{at} \cdot f_{at}, \tag{32}$$

where u_{ca}, p_{at} and f_{at} are as defined above.

Using $\log \widetilde{H}_{ct}$ instead of $\log M_{ct}$ in column 7 leads to a similar coefficient relative to column 1 of Table 3A (4.27 with standard error of 1.24). The difference in magnitude may be because

insurance-driven and demographics-driven market sizes affect different types of drugs, or because the demographics-driven changes are anticipated further in advance, potentially enabling a greater response. The most likely explanation, however, is that $\widetilde{H}_{ct}$ is a worse measure of market size than M_{ct} because (32) effectively assigns 0 use to those without insurance.[35]

We next construct an alternative measure of $\widetilde{H}_{ct}$ from information on Medicaid coverage. Changes in Medicaid eligibility have made children more likely to be covered. In the 1970s, about 9% of 0-20 year olds were covered by Medicaid; this fraction rose to 17% by the late 1990s. Column 8 shows a very small and marginally significant effect of Medicaid insurance on the rate of entry of new drugs.

Finally, as a falsification exercise, we use a measure of $\widetilde{H}_{ct}$ calculated from information on Medicare take-up rates. Since Medicare does not cover prescription drugs, this measure should not predict new drug entries (though note that M_{ct} and $\widetilde{H}_{ct}$ are correlated by construction). Reassuringly, the estimate in column 9 shows no positive effect of Medicare coverage.

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 drug-induced changes in population are likely to be small relative to the demographic changes that we are exploiting. Nevertheless, we further 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 are developed in the intervening 10 years.

These instrumental-variables (IV) estimates show no evidence of reverse causality. In column 2 of Table 7, we instrument for market size with past market size. With NLLS, the estimate is 5.56 (s.e. =1.69), 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 IV estimates in column 4 give larger coefficients than the corresponding non-instrumented results, particularly with NLLS, and both are significant at the 1 percent level.

[35] Another possibility is that the effect of demographics-driven market size was somewhat overestimated in Table 3A because of the correlation between this measure and insurance coverage. However, the estimates in columns 1-6 in this table suggest that this is not a major concern.

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 because the NAMCS has a more even distribution of users across drug categories than the MEPS, and also because it starts in 1980, enabling us to check whether use of drug consumption and expenditure data from the late 1990s introduces any biases. However, as noted above, the NAMCS is less representative, since the data are reported by doctors in private practice, whereas the MEPS is a sample of all U.S. households.

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, 2.86, is insignificant and considerably smaller than the corresponding MEPS estimate. The OLS estimate is also smaller than the corresponding MEPS estimate, but is significant at the 1 percent level. Column 6 uses the MEPS data with the NAMCS classification system, and obtains estimates similar to column 3 of Table 3A. This shows that the disparity between the NAMCS results and Table 3A is not due to differences in classification systems, but is probably driven by the non-representative nature of the NAMCS data.[36]

In column 7, we use ten-year time intervals, and find similar results to column 5; the NLLS is insignificant while the OLS is significant at the 5 percent level. Columns 8 and 9 report unweighted results, which are larger than the weighted estimates. This contrasts with the pattern from the MEPS, and is consistent with our conjecture that the differences between weighted and unweighted results in the MEPS were largely because of the less precise age distribution estimates for the smaller categories in that data set (this is not a problem in the NAMCS, which leads to a more precise estimate 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.[37] We construct an alternative estimate of market size, M_{ct}^{1980}, with the use per person numbers, u_{ca}^{1980}, only from the 1980 NAMCS survey. We then

[36] Using an alternative market size variable constructed with the fraction of total use for each age group for u_{ca}, rather than use per person for each age group, leads to very similar estimates from the two surveys, which are also close to the MEPS estimates reported in Table 3A. Although this alternative measure has the advantage of not being sensitive to the total expenditure by category in NAMCS, the baseline measure we use appears to be both more natural and theoretically better motivated.

[37] 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 shares 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 importance of Gastrointestinals for 30-50 year-olds. Nevertheless, this will not lead to spurious positive estimates of the effect of potential market size on entry, since overtime variation in our measure of market size is still purely driven by demographic changes.

estimate equations (24) and (26) in the post-1980 sample, instrumenting $\log M_{ct}$ with $\log M_{ct}^{1980}$. For comparison, using our baseline measure only for 1980 onwards with the NAMCS data, we obtain a NLLS estimate of 4.64 (s.e.= 3.58) and an OLS estimate of 5.76 (s.e.= 4.00) in column 10. Instrumenting with $\log M_{ct}^{1980}$ in column 11 leads to an estimate of α equal to 4.79 (s.e.= 3.98) with NLLS and 6.63 (s.e.= 4.57) with OLS, in both cases similar to the estimate in column 10. This comparison suggests that using microdata from the 1990s to construct rates of drug use is unlikely to create any significant bias.

5.7 Generics vs Non-Generics

Table 8 shows the results of some of our specifications using only generic drugs to construct N_{ct}. Table 9 reports similar regressions for non-generics. Because we now look at generics and non-generics separately, there are more empty cells; for example, using all drugs and five-year intervals, there are 9 empty cells, but there are 27 and 24 for generics and non-generics, respectively. Therefore, in Tables 8 and 9, the NLLS results should be more reliable than the OLS estimates (but we also report the latter for completeness).

Column 1 in both Tables 8 and 9 reports the basic specifications for five-year intervals using the income-based measure of market size, similar to column 1 in Table 3A. For generics, the NLLS estimate, 7.70 (s.e.=2.69), is significant at the 1 percent level.[38] The corresponding estimate in Table 9 for non-generics, 4.41 (s.e.=1.91), is smaller, though still significant at 1 percent. 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.

Columns 2 and 3 investigate the robustness of this result when we control for serial correlation by including lagged entry on the right hand side and when we control for major drug category trends (respectively, estimating equations (28) and (30)). The lagged dependent variable is not significant and has little effect on the point estimates of α either with generics or non-generics, though with generics, the estimate, 8.08 (s.e.= 5.34), is only significant at 10 percent. The estimate for non-generics, 4.55 (s.e.= 1.91), continues to be significant at 1 percent. Controlling for linear trends, on the other hand, has almost no effect on the estimates, which continue to be significant at 1 percent in both cases.

Columns 4 and 5 investigate anticipation effects. For generics, when both current and lead market sizes are entered together, neither market size is significant, though the estimate on

[38] Because the requirements to gain FDA approval for generic drugs changed in 1984 with the Hatch-Waxman Act, we investigated whether the relationship between potential market size and new drug approvals also changed. The answer seems to be no. For example, interacting market size with a post-1985 dummy in the NLLS specification, the estimate of α is 7.94 (s.e. 2.71) and the coefficient on the interaction is 0.15 (s.e. 0.08), i.e., quantitatively very small. There is a large increase in approvals of generics in the late 1980s followed by a decrease in the early 1990s, which are mostly captured by the time dummies.

current market size is similar to column 1, while the estimate on lead market size is negative and small. When lead market size is entered by itself, it is significant for five-year intervals, thus there might be some limited response of generics to anticipated future market sizes. The pattern is somewhat different for non-generics: lead market size is insignificant both when entered together with current market size and when used by itself.[39]

Columns 6 and 7 of Table 8 investigate the impact of lag market size. When current and lag market size are present together, current market size is significant and lag market size has a negative and insignificant coefficient. On the other hand, lag market size on its own has a statistically significant effect on entry of generics in column 7. The evidence is therefore inconclusive as to whether lag market size matters for generic entry.[40] In contrast, in non-generics regressions lag market size is insignificant both when entered together with current market size and by itself, and we do not report these specifications in Table 9 to save space.

The FDA has labeled some non-generics as priority drugs because they appear to offer significant clinical improvement over available products and therapies. The FDA has a second classification, whether or not a drug contains a new molecule (an active ingredient that has not been marketed before in the U.S.). Research by the National Institute for Health Care Management (2002) suggests that drugs labeled priority or new molecules (but not both) generate less revenue than other non-generics, but those that are both priority drugs and new molecules generate more revenues. To look at the relationship between market size and a potential measure of more significant innovation, we construct a new measure of N_{ct}, which only includes approvals for drugs that are designated to be both new molecular entities and priority drugs. In the basic specification reported in column 6 of Table 9, we find an estimate of 4.06, which is very similar to the estimate in column 1 but statistically insignificant because of the large standard error (3.48). Column 7 looks at the impact of the five-year lead market size, and obtains a coefficient estimate of 4.71 (s.e.= 2.79), which is significant at 10 percent. Therefore, these estimates are similar to the estimates from columns 1 and 2, but the standard errors are considerably larger. Most probably, the significantly larger standard errors result from the presence of many fewer new molecular entities and priority drugs than generics or non-generics in our data (there are only a total of 190 such drugs in our 1970-2000 dataset).

[39] We also experimented with longer leads, and obtained similar results. Lack of a significant anticipation effect for non-generics is somewhat surprising, especially given the evidence that there might be some anticipation effects for generics. This might be because the non-generic results are less precise, making it difficult to detect anticipation effects. Note also that when we look at priority drugs and new molecules in columns 6 and 7, lead market size has more predictive power than current market size, but is only significant at 10 percent.

[40] We expected lag market size to matter and generic entry to lag behind non-generics because of patent protection for non-generics.

5.8 Patents

Finally, we investigate the effect of potential market size on patents. We obtained data on pharmaceutical patents from Thomson Derwent Inc., and with the help of a specialist at this company, we mapped these patents into our FDA classification system.[41] However, the mapping of patents for chemicals into FDA categories is imperfect and necessarily introduces a significant amount of noise, which makes inference more difficult in this case.

Firms typically apply for a patent prior to the clinical trial stage of drug development, or about 5-10 years before the drug is approved.[42] Given the results so far, we might expect patents to respond to future demographic changes.

In columns 8 and 9 of Table 9, we estimate equations (24) and (26), except that N_{ct} is now the number of patents in a particular drug category. The estimates of α are positive and of similar magnitude to the estimates for non-generics (e.g., 3.12 for current market size and 4.38 for lead market size), but because of the much larger standard errors, they are statistically insignificant.[43] There may be a number of reasons for the much larger standard errors with patents.[44] First, this may simply reflect the imperfect match between the patent data and the FDA categories, especially bearing in mind the potential use of certain chemical structures in multiple drug lines. Second, the significant costs and uncertainty involved in the development of new molecules and patentable products may be creating substantial attenuation (e.g., a drug intended for the 1990s may be patented in the 1980s or 1990s, depending on delays in the research process). Third, pharmaceutical companies may respond more to profit incentives at the later stages of the research process than at the earlier stages. Finally, patents may be more responsive to OECD demand than to U.S. demand. To investigate this last possibility, we looked at the relationship between changes in market size derived from European, Japanese and U.S. demographic changes. In this case, we find estimates of the five-year lead of market size on patents that are broadly similar to the results for non-generics and statistically significant at 5 percent (for example, 4.23 with standard error 1.82 with NLLS and 3.07 with standard error 1.28 with OLS). Although this result suggests that OECD demand may be more important for patents, we are currently unable to make more progress in distinguishing between these various

[41] We could not use the data from the Hall-Jaffe-Trachtenberg patent dataset (see Jaffe and Trachtenberg, 2002) because we were unable to map their classification based on chemical structure to our drug categories.

[42] The firm therefore loses a significant fraction of the life of the patent before it can begin marketing the drug. Part of the Hatch-Waxman Act allowed pharmaceutical companies to apply to the FDA for an extension of the life of their patents, if they could show that they lost marketing time while waiting for approval. The maximum extension is 5 years, and depends, among other things, on the length of the initial FDA approval process. Overall, companies have a maximum of 14 years of patent protection after FDA approval.

[43] We obtain similar results with ten-year or fifteen-year leads of market size.

[44] Finkelstein (2003) also finds weaker results for vaccine patents than for later stages of development.

explanations, and the weaker results for patents remain a puzzle.

6 Concluding Remarks

This paper investigates the response of entry of new drugs and pharmaceutical innovation to changes in potential market size of users, driven by U.S. (or OECD) demographic changes. Our results indicate that a 1 percent increase in the potential market size for a drug category leads to approximately 4-6 percent growth in the entry of new drugs approved by the FDA. This response comes from the entry of both generics and non-generics, though the effect on generics is larger and somewhat more 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 more profitable areas. Furthermore, the magnitude of the effect, which is important for evaluating various theoretical predictions of these models, is substantial. For example, directed technical change models suggest that the relative demand curves for factors can be upward, rather than downward, sloping if the 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)—the corresponding number implied by our estimates is in the range of 4 to 6. 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 may be nonrepresentative, for example because the pharmaceuticals may be more research oriented than other industries. Future research investigating the response of innovation and entry of new products to market size both in specific industries and at the economy-wide level is necessary to substantiate the results presented here.

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*, 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.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning-by-Doing," *Review of Economic Studies*, 29, pp. 155-173.

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, pp. 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.

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.

DiMasi, Joseph A., Ronald W. Hansen and Henry G. Grabowski. 2003. "The Price of Innovation: New Estimates of Drug Development Costs." *Journal of Health Economics*, 22, pp. 151-185.

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

Finkelstein, Amy. 2003. "Health Policy and Technical Change: Evidence from the Vaccine Industry." *NBER Working Paper* #9460.

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.

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, pp. 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. 1996. "Scale, Scope and Spillovers: The Determinants of Research Productivity in Drug Discovery." *Rand Journal of Economics,* 27, pp. 32-59.

IMS. 2000. "World Review" IMS Health, at www.ims-global.com/insight/report/market_growth/repor

Jaffe, Adam and Manuel Trachtenberg and Rebecca Henderson. 2002. *Patents, Citations, and Innovations.* MA: MIT Press..

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. 2001. "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. 27-52.

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". Columbia University 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 Technologies, 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.

National Institute for Health Care Management. 2002. "Changing Patterns of Pharmaceutical Innovation" at www.nihcm.org/innovations.pdf.

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.

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

Reiffen, David and Michael R.Ward. 2002 "Generic Drug Industry Dynamics." University of Texas at Arlington, mimeo

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, pp. 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

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, 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 Table 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 national drug code 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 results are very similar if we construct use per person as a weighted average of use per person in the survey, i.e., without using CPS information. We prefer the former methodology because it enables us to construct income-based measures without relying on MEPS income data.

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 age profile of users, 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, we aggregate the initial ten-year age groups into 5 age groups, 0-20, 20-30, 30-50, 50-60, 60+, since 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, 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.[45] 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 is in 1980, where we matched about 85% of prescribed drugs; in most years the match rate is 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 1990s 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, which is available upon request. We find that the same 5 age groups are suitable for the NAMCS data. We use the same cutoff rule of 1500 observations for the NAMCS as for the MEPS for dropping FDA categories. Different FDA categories are excluded from the two surveys, e.g., Antifungals from the NAMCS 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. Over-the-counter drugs and orphan drugs (of which only a few can be matched) are excluded. Moreover, biologics, which go through a separate approval process, are not in this dataset.

We match drugs in the approval list to FDA categories by drug name. 13,916 of 16,220 prescription drugs (86%) approved since 1970 are matched, while before 1970, the match rate is about 45%. This motivates our focus on drug approvals between 1970 and 2000. Drugs that have the same approval number and FDA class as a previously approved drug and drugs for which the corresponding FDA category is dropped because of insufficient observations in the MEPS are excluded. Finally, we drop drugs with the same name, MEPS category and different dosage from a previously approved drug, leaving us with our sample of 6,595 drugs. Of these, 1,928 are non-generics, and 190 are priority drugs and new molecules.

[45] 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.

8.4 CRISP Data

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 earliest 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 has about 2,900 diseases, though these are arranged in a heirarchical structure into 35 major disease classes. We map the detailed disease classes into our classification scheme, though there are no matches of primary disease to the five 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.

8.5 Patents

We have obtained patent data from Thomson Derwent Inc. We use all pharmaceutical patents granted in the United States, between 1970-2000. We use these data instead of the Hall-Jaffe-Trachtenberg patent data because the latter use a classification for pharmaceuticals based on chemical structure, which we are not able to map into our FDA classification system. The Thomson Derwent patents are classified by chemical structure and therapeutic intent, and a specialist at the company has mapped this system into the FDA system. Because the mapping is not precisely one to one, about five percent of the patents fall into two of our categories, which we drop. We are left with 275,406 patents. This number is comparable to the number of pharmaceutical patents in the Hall-Jaffe-Trachtenberg patent dataset. Since we were unable to obtain citations to construct a weighting system, we assign all patents equal weight.

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	0.851	0.778	0.732

Panel B: MEPS over time

	1996/1997	1997/1998	1996/1998
Overall Correlation	0.996	0.996	0.992
Weighted Correlation	0.999	0.999	0.997
Mean Correlation by Drug	0.919	0.955	0.902

Panel C: Correlation Between NAMCS and MEPS, and Between MEPS Use and Expenditure

	MEPS/NAMCS	MEPS use/MEPS expenditure
Overall Correlation	0.187	0.962
Weighted Correlation	0.360	0.983
Mean Correlation by Drug	0.813	0.853

Notes: Overall correlation is the correlation of use per person (or average expenditure share for expenditure cells) across all categories. In weighted correlations, observations are weighted by total use or expenditure 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	26.29	26.74	27.43
20-30	25.96	26.61	27.09
30-50	26.32	27.14	28.04
50-60	25.78	26.31	27.11
60+	25.82	26.58	27.30
Panel B: Log Drug Approvals of All Drugs by Age Group, From FDA			
Age Group	1970-1980	1980-1990	1990-2000
0-20	6.49	6.33	5.37
20-30	3.81	4.54	3.87
30-50	6.22	6.85	6.23
50-60	5.57	6.22	5.86
60+	6.22	7.21	6.38
Panel C: Log Drug Approvals of Generics by Age Group, From FDA			
Age Group	1970-1980	1980-1990	1990-2000
0-20	6.32	5.90	3.95
20-30	2.64	3.74	2.89
30-50	6.06	6.68	5.68
50-60	5.33	5.88	5.36
60+	5.92	6.96	5.81
Panel D: Log Drug Approvals of Non-generics by Age Group, From FDA			
Age Group	1970-1980	1980-1990	1990-2000
0-20	4.62	5.29	5.09
20-30	3.43	3.95	3.40
30-50	4.34	4.96	5.37
50-60	4.03	4.98	4.91
60+	4.90	5.71	5.53

Notes: In Panel A, log income by age group is the log of mean income over the ten year interval, from the March CPS. In panels B, C and D each of the 34 drug categories is assigned one age category, based on the age group that uses that category most. See text and Appendix Table A1 for details. Log of drug approvals is the log of the sum of all approvals in the indicated 10 year interval, for all drug categories corresponding to 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	6.04 (1.95)	4.95 (1.90)	6.16 (2.04)	5.16 (1.84)	5.51 (1.84)	4.02 (1.51)
R Squared	0.86	0.91	0.86	0.91	0.75	0.84
	Panel B: OLS, dependent variable is log drug approvals					
Log Market Size	4.52 (2.00)	4.65 (2.13)	5.74 (2.42)	5.27 (2.21)	3.41 (1.62)	4.39 (2.11)
R Squared	0.87	0.92	0.86	0.91	0.80	0.86
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: Huber-White robust standard errors are reported in parentheses. 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 average expenditure share 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. 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). See equation (26) in the text. 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; see equation (24). 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	6.25 (0.52) {1.53}	5.98 (0.57) {1.50}	6.55 (0.67) {1.81}	6.23 (0.74) {1.73}
	Panel B: Weighted Poisson ML, dependent variable is count of drug approvals			
Log Market Size	6.56 {1.74}	6.18 {1.91}	8.58 {1.66}	7.90 {1.70}
	Panel C: Negative Binomial ML, dependent variable is count of drug approvals			
Log Market Size	5.40 (1.58) {2.07}	5.11 (1.70) {1.64}	5.39 (2.14) {2.87}	5.27 (2.30) {2.25}
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: Maximum Likelihood standard errors in parentheses, and Huber-White standard errors in curly brackets. Drug approvals, 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 the Poisson model is estimated using maximum likelihood. In panel C the negative binomial model is estimated using maximum likelihood.

Table 4:

Delays and Anticipation Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
				Panel A: NLLS, dependent variable is count of drug approvals						
Log Market Size	6.04	4.95	13.56	6.50			1.06	5.04		
	(1.95)	(1.90)	(4.45)	(2.24)			(3.97)	(1.93)		
Lag Market Size			-6.16	-3.89	3.34	-0.39				
			(3.15)	(2.20)	(1.38)	(2.24)				
Lead Market Size							7.64	5.01	8.78	8.69
							(4.86)	(3.65)	(2.30)	(3.13)
R Squared	0.86	0.91	0.86	0.91	0.86	0.90	0.88	0.96	0.88	0.95
				Panel B: OLS, dependent variable is log drug approvals						
Log Market Size	4.52	4.65	8.62	5.67			0.81	5.41		
	(2.00)	2.13	(3.91)	(2.65)			(3.35)	(2.80)		
Lag Market Size			-3.32	-2.14	2.29	1.34				
			(2.97)	(2.43)	(1.47)	(2.15)				
Lead Market Size							6.00	4.92	6.84	8.08
							(4.39)	(5.23)	(2.76)	(4.67)
R Squared	0.87	0.92	0.87	0.91	0.86	0.91	0.88	0.96	0.88	0.96
Number of Observations	204	102	204	102	204	102	170	68	170	68
Length of Time Interval (Years)	5	10	5	10	5	10	5	10	5	10

Notes: Huber-White standard errors in parentheses. 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.

Table 5:

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	5.78 (1.89)	6.44 (1.75)	5.28 (1.76)	5.24 (1.83)	5.74 (1.84)	5.72 (1.83)	6.23 (1.88)	6.24 (1.72)	5.60 (2.10)	5.43 (1.88)	5.93 (1.99)	6.32 (2.11)
Lagged Dependent Variable	0.02 (0.29)	0.52 (0.35)										
R Squared	0.86	0.92	0.85	0.85	0.86	0.86	0.57	0.61	0.49	0.83	0.85	0.85
			Panel B: Linear Regressions, dependent variable is log drug approvals									
Log Market Size	4.23 (1.99)	5.16 (2.21)	5.00 (2.18)	4.73 (2.22)	4.46 (1.97)	4.68 (2.26)	4.66 (2.17)	4.68 (1.97)	3.23 (3.33)	4.25 (2.09)	4.55 (2.00)	4.66 (2.04)
Lagged Dependent Variable	-0.43 (0.42)	1.98 (0.68)										
R Squared	0.87	0.93	0.88	0.88	0.87	0.89	0.87	0.87	0.89	0.77	0.87	0.86
			Panel C: NLLS for Poisson model, dependent variable is count of drug approvals									
Lead Market Size			8.78 (2.30)	8.77 (2.50)	8.78 (2.47)	8.63 (2.43)	9.03 (2.51)	8.36 (2.45)	8.46 (2.53)	8.62 (2.66)	8.76 (2.51)	8.77 (2.47)
R Squared			0.88	0.88	0.88	0.89	0.89	0.89	0.62	0.86	0.88	0.88
Length of Time Interval (Years)	5	10	5	5	5	5	5	5	5	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 Trends	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	Thyroid, Blood	Biologics Included

Notes to Table 5: Huber-White standard errors in parentheses. 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. 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, and 12; 136 in column 6; 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 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 one-half 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 for the 16 major drug categories. See text for details. In column 12 FDA approvals of biologics for each category and time interval are added to the dependent variable.

Table 6:

Controlling for Changes in Health Insurance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: NLLS for Poisson model, dependent variable is count of drug approvals								
Log Market Size	5.21	5.79	5.85				4.27	0.19	-3.05
	(2.30)	(1.92)	(2.04)				(1.24)	(0.12)	(3.36)
Lead Market Size				7.79	8.10	8.51			
				(2.72)	(3.35)	(2.47)			
Health Insurance Market Size	2.59	0.08	-1.68	4.16	-0.35	-5.80			
	(2.93)	(0.12)	(3.37)	(3.77)	(1.23)	(4.74)			
R Squared	0.86	0.86	0.86	0.88	0.88	0.88	0.86	0.85	0.85
	Panel B: OLS, dependent variable is log drug approvals								
Log Market Size	3.58	3.98	4.32				3.11	0.25	-2.22
	(2.22)	(1.99)	(2.04)				(1.21)	(0.08)	(2.94)
Lead Market Size				5.89	6.25	6.65			
				(2.94)	(3.29)	(2.77)			
Health Insurance Market Size	2.36	0.20	-1.70	3.48	-0.30	-4.85			
	(2.38)	(0.08)	(2.78)	(3.48)	(1.08)	(4.90)			
R Squared	0.87	0.87	0.87	0.88	0.88	0.88	0.87	0.86	0.86
Number of Observations	204	204	204	170	170	170	204	204	204
Length of Time Interval (Years)	5	5	5	5	5	5	5	5	5
Type of Insurance	Any Private	Medicaid	Medicare	Any Private	Medicaid	Medicare	Any Private	Medicaid	Medicare

Notes: Huber-White standard errors in parentheses. Drug approvals and market size variables are constructed as in Table 3A. In columns 1-6 log market size and lead market size are computed as in Table 3A. Health insurance market size is computed by multiplying the time-invariant average expenditure share of users in a particular age group, by average income of that age group for 1970-2000, by the fraction of people in that age group with with the corresponding type of health insurance. In columns 7-9 market size is obtained by multiplying the time-invariant average expenditure share 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.

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	6.04	5.56	4.95	8.54	2.86	6.18	2.90	4.41	4.60	4.64	4.79
	(1.95)	(1.69)	(1.90)	(3.37)	(2.04)	(2.59)	(3.73)	(2.49)	(2.51)	(3.58)	(3.98)
R Squared	0.86	0.86	0.91	0.91	0.76	0.83	0.83	0.68	0.78	0.63	0.63
	Panel B: Linear Regressions, dependent variable is log of drug approvals										
Log Market Size	4.52	3.91	4.65	5.96	3.76	6.51	3.96	4.84	6.45	5.76	6.63
	(2.00)	(2.11)	(2.13)	(3.94)	(1.39)	(2.84)	(1.95)	(1.61)	(3.20)	(4.00)	(4.57)
R Squared	0.87	0.87	0.92	0.92	0.86	0.83	0.93	0.82	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
Data Set Used for Market Size	MEPS	MEPS	MEPS	MEPS	NAMCS	MEPS	NAMCS	NAMCS	NAMCS	NAMCS	NAMCS
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: Huber-White standard errors in parentheses. 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. In columns 5-11, regressions include 30 drug categories, constructed from the NAMCS, 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 are used. In column 11 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)
			Panel A: NLLS, dependent variable is count of drug approvals				
Log Market Size	7.70 (2.69)	8.55 (5.34)	8.08 (2.12)	8.44 (5.39)		14.54 (7.27)	
Lagged Dependent Variable		0.84 (0.57)					
Lead Market Size				-1.81 (7.35)	7.64 (3.24)		
Lag Market Size						-5.26 (5.16)	4.90 (1.88)
R Squared	0.78	0.72	0.86	0.79	0.79	0.79	0.78
			Panel B: OLS, dependent variable is log drug approvals				
Log Market Size	9.03 (2.42)	12.80 (6.39)	8.01 (3.18)	5.68 (4.45)		13.36 (6.84)	
Lagged Dependent Variable		1.81 (1.01)					
Lead Market Size				4.10 (6.91)	10.02 (3.69)		
Lag Market Size						-3.48 (4.77)	5.15 (1.62)
R Squared	0.84	0.90	0.86	0.85	0.85	0.84	0.84
Number of Observations	204	136	204	170	170	204	204

Notes: Hubler-White standardard errors in parentheses. 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. All time periods are 5 years. Lagged dependent variable is instrumented with twice lagged dependent variable, as in Table 6.

Table 9:

Non - Generics and Patents: Effect of Changes in Market Size on Innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Panel A: NLLS, dependent variable is count of drug approvals or patents					
Log Market Size	4.41 (1.91)	4.55 (1.91)	4.49 (1.69)	-3.37 (3.30)		4.06 (3.48)		3.12 (2.32)	
Lagged Dependent Variable		-0.11 (0.17)							
One Period Lead Market Size				5.99 (4.10)	2.61 (2.28)		4.71 (2.79)		4.38 (3.10)
R Squared	0.83	0.85	0.85	0.85	0.85	0.83	0.80	0.88	0.94
				Panel B: OLS, dependent variable is log drug approvals or patents					
Log Market Size	1.97 (1.78)	3.50 (2.00)	2.29 (1.90)	-1.01 (2.99)		1.03 (1.43)		2.27 (1.47)	
Lagged Dependent Variable		-0.69 (0.67)							
One Period Lead Market Size				1.93 (4.05)	0.89 (2.25)		-0.70 (2.07)		2.52 (1.76)
R Squared	0.84	0.86	0.88	0.86	0.88	0.87	0.84	0.94	0.97
Dependent Variable	All Non-Generics	All Non-Generics	All Non-Generics	All Non-Generics	All Non-Generics	Priority & New Mol.	Priority & New Mol.	Patents	Patents
Number of Observations	204	170	170	204	204	204	102	204	170

Notes: Huber-White standard errors in parentheses. Dependent variables in columns 1-5 include only approvals of non-generic drugs. Dependent variables in columns 6 and 7 include drugs that the FDA has designated as both priority and new mocecules, as described in the Appendix. Patents counts for the dependent variable in columns 8 and 9 are computed from the Derwent Inc. patent dataset, by counting patents for each category over ten-year intervals (see Appendix for details). 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. All time periods are 5 years. Lagged dependent variable is instrumented with twice lagged dependent variable, as in Table 6.

Appendix Table A1:

Summary of Classifications and Drug Use by Age Group, Computed From MEPS, 1996-1998

Class	Description	Use Per Person (Share of Use in Parentheses) 0-20	20-30	30-50	50-60	60+	Age Group with Peak Share of Use
10	Penicillins, Cephalosporins, Lincosamides, Sulfonamides, Misc. Antibacterials	0.61 (0.40)	0.30 (0.09)	0.38 (0.26)	0.44 (0.10)	0.45 (0.16)	0-20
11	Tetracyclines, Urinary Tract Antiseptics, Quinolones	0.02 (0.11)	0.06 (0.13)	0.06 (0.31)	0.09 (0.14)	0.12 (0.30)	30-50
12	Antifungals, Antivirals	0.03 (0.16)	0.03 (0.08)	0.08 (0.43)	0.09 (0.15)	0.07 (0.18)	30-50
20	Hematologics	0.00 (0.01)	0.00 (0.01)	0.03 (0.11)	0.11 (0.12)	0.43 (0.75)	60+
30	Cardiovascular - Renal	0.05 (0.01)	0.10 (0.01)	0.69 (0.15)	2.68 (0.19)	6.05 (0.65)	60+
40	Sedatives/Hypnotics, Antianxiety	0.01 (0.02)	0.04 (0.04)	0.16 (0.23)	0.27 (0.38)	0.41 (0.33)	50-60
41	Antipsychotics/Antimania, Antidepressants	0.08 (0.06)	0.19 (0.07)	0.57 (0.46)	0.70 (0.18)	0.57 (0.23)	30-50
42	Anorexiants	0.04 (0.46)	0.01 (0.07)	0.03 (0.35)	0.02 (0.09)	0.01 (0.04)	0-20
43	Misc. Central Nervous System	0.11 (0.75)	0.01 (0.04)	0.01 (0.09)	0.01 (0.03)	0.02 (0.08)	0-20
50	Gastrointestinals	0.03 (0.03)	0.08 (0.04)	0.24 (0.27)	0.52 (0.19)	0.83 (0.47)	60+
60	Hyperlipidemia, Electrolyte Replenishment/Regulation, Calcium Metabolism	0.01 (0.01)	0.01 (0.00)	0.13 (0.12)	0.67 (0.21)	1.37 (0.66)	60+
61	Vitamins/Minerals	0.06 (0.20)	0.09 (0.16)	0.07 (0.27)	0.09 (0.12)	0.13 (0.25)	60+
70	Adrenal Corticosteroids	0.05 (0.13)	0.04 (0.06)	0.09 (0.29)	0.14 (0.14)	0.24 (0.38)	60+
71	Androgens/Anabolic Steroids, Estrogens/Progestins	0.02 (0.02)	0.38 (0.13)	0.34 (0.26)	1.30 (0.33)	0.67 (0.26)	50-60
72	Blood Glucose, Thyroid	0.02 (0.01)	0.10 (0.03)	0.35 (0.22)	1.02 (0.21)	1.70 (0.54)	60+
73	Contraceptives	0.01 (0.06)	0.21 (0.47)	0.10 (0.42)	0.01 (0.02)	0.01 (0.03)	20-30
80	Immunologics	0.00 (0.08)	0.01 (0.09)	0.02 (0.33)	0.02 (0.14)	0.03 (0.36)	60+

Appedix Table A1 (cont.)

Class	Description	Use Per Person (Share of Use in Parentheses) 0-20	20-30	30-50	50-60	60+	Age Group with Peak Share of Use
90	Dermatologics, Topical Anti-Infectives	0.09 (0.25)	0.09 (0.26)	0.09 (0.27)	0.12 (0.12)	0.17 (0.11)	20-30
91	Topical Steroids	0.01 (0.18)	0.01 (0.07)	0.01 (0.27)	0.02 (0.11)	0.04 (0.37)	60+
100	Extrapyramidal Movement	0.00 (0.03)	0.01 (0.03)	0.03 (0.34)	0.03 (0.11)	0.08 (0.49)	60+
101	Skeletal Muscle Hyperactivity, Anticonvulsants	0.05 (0.08)	0.11 (0.08)	0.26 (0.44)	0.29 (0.16)	0.27 (0.23)	30-50
110	Oncolytics	0.00 (0.02)	0.01 (0.03)	0.05 (0.25)	0.16 (0.27)	0.17 (0.44)	60+
120	Misc. Ophthalmics, Glaucoma	0.00 (0.00)	0.00 (0.01)	0.01 (0.05)	0.06 (0.08)	0.41 (0.86)	60+
121	Ocular Anti-Infective	0.06 (0.23)	0.05 (0.09)	0.05 (0.22)	0.08 (0.11)	0.16 (0.36)	60+
130	Topical Otics	0.02 (0.30)	0.01 (0.07)	0.01 (0.22)	0.03 (0.14)	0.04 (0.27)	0-20
131	Vertigo/Motion Sickness	0.02 (0.12)	0.01 (0.05)	0.03 (0.24)	0.05 (0.13)	0.13 (0.47)	60+
140	General Analgesics, Narcotic Analgesics, Antiarthritics, NSAID	0.09 (0.05)	0.29 (0.08)	0.59 (0.35)	0.89 (0.17)	1.13 (0.35)	30-50
141	Non-Narcotic Analgesics	0.00 (0.06)	0.01 (0.03)	0.02 (0.32)	0.04 (0.17)	0.07 (0.43)	60+
142	Antigout	0.00 (0.00)	0.00 (0.00)	0.01 (0.13)	0.06 (0.19)	0.14 (0.68)	60+
143	Central Pain Syndromes	0.01 (0.15)	0.01 (0.10)	0.02 (0.45)	0.02 (0.15)	0.01 (0.15)	30-50
150	Antiparasitics	0.00 (0.07)	0.01 (0.07)	0.03 (0.37)	0.03 (0.12)	0.05 (0.37)	30-50
160	Antiasthmatics, Nasal Decongestants	0.20 (0.20)	0.14 (0.07)	0.22 (0.23)	0.47 (0.16)	0.66 (0.35)	60+
161	Antitussives, Antihistamines, Corticosteroids	0.15 (0.17)	0.20 (0.10)	0.29 (0.33)	0.45 (0.17)	0.41 (0.24)	30-50
162	Cold Remedies	0.07 (0.29)	0.05 (0.09)	0.07 (0.31)	0.08 (0.12)	0.08 (0.18)	30-50

Notes: Construction of the 34 categories is described in the Data Appendix. Each category includes the indicated FDA sub-categories. Use per person is the mean number of drugs in the class used per person in the age group. Share of use is the fraction of drugs used in the category by the age group.

Figure 1: Response to an Anticipated Increase in Market Size

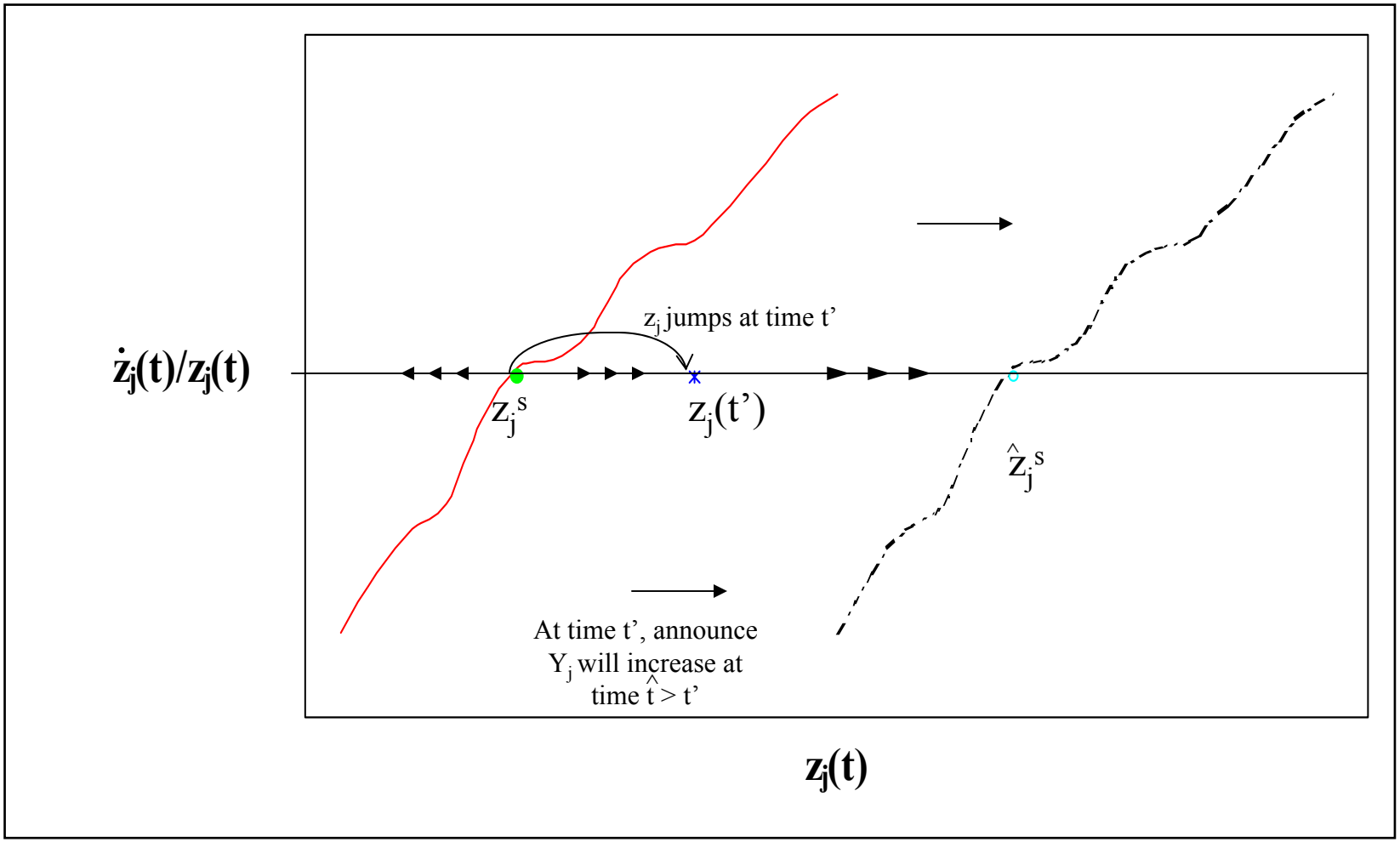


Figure 2: Share of Population by Age Group, 1964-2000, from CPS

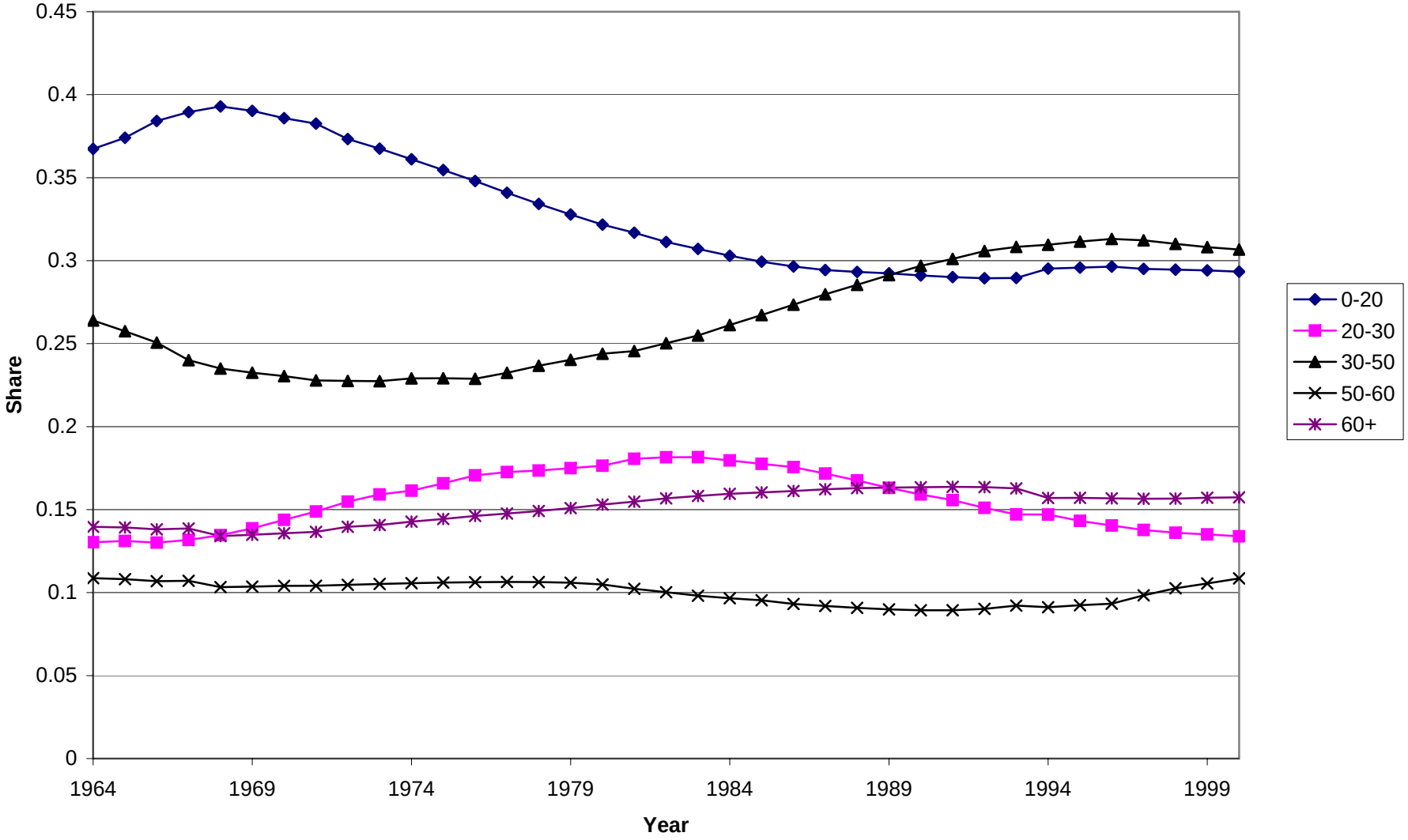


Figure 3: Share of Income by Age Group, 1970-2000, from CPS

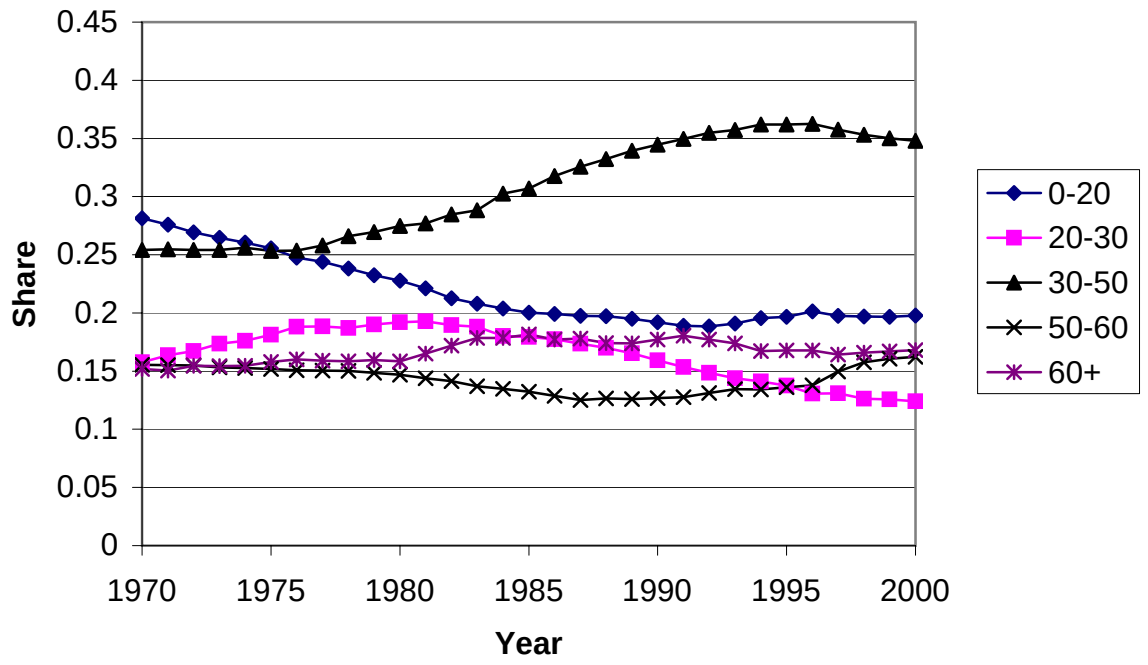
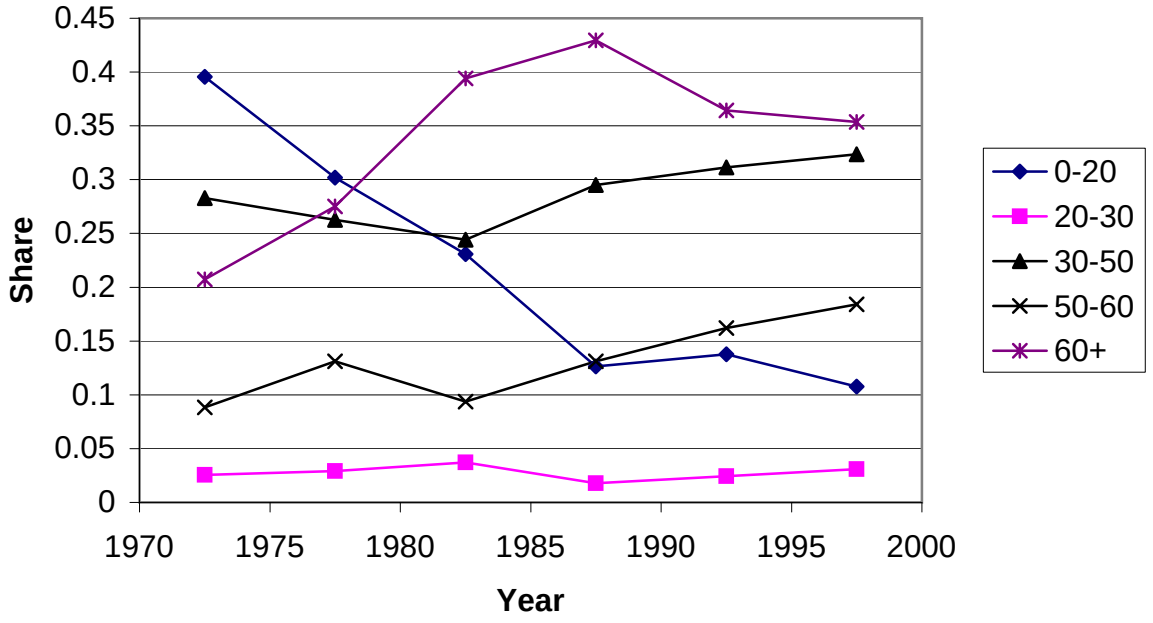
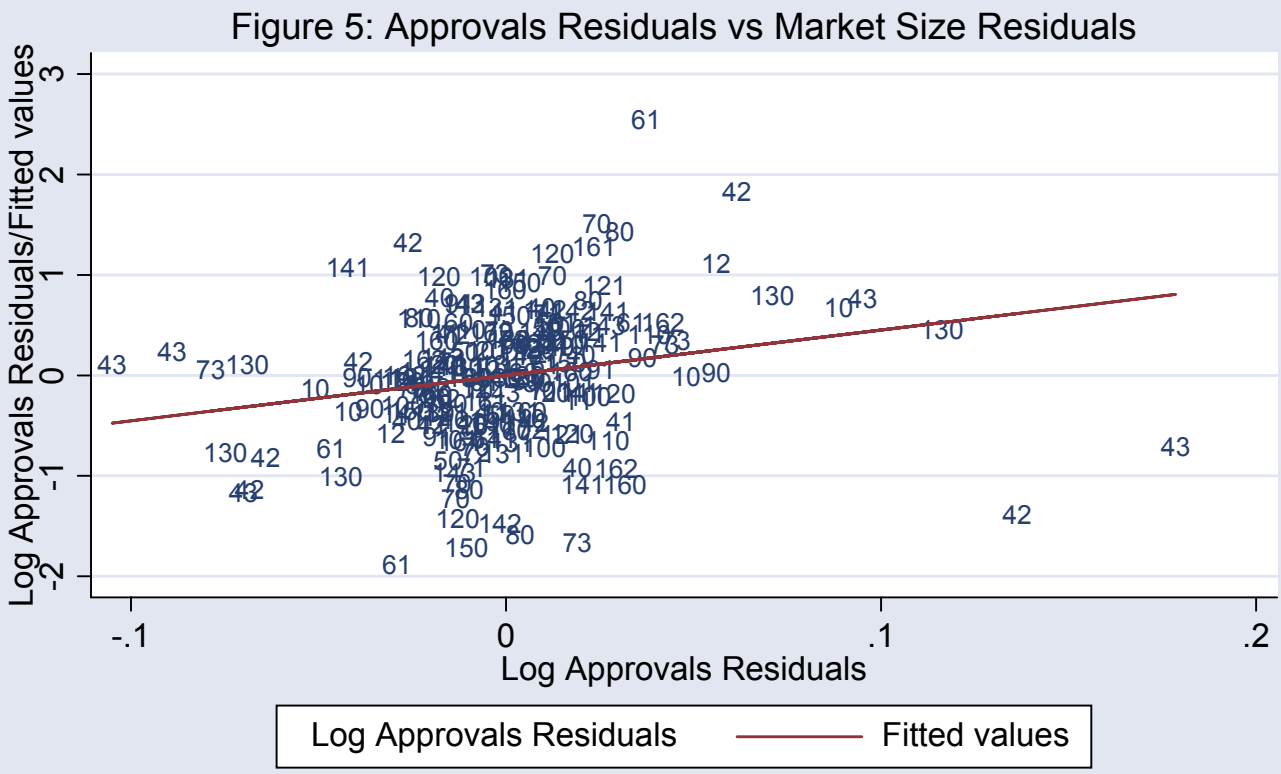


Figure 4: Share of Drug Approvals by Age Group, 1970-2000, from FDA





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